

T9519

STIC-Biotech/ChemLib

From: Slobodyansky, Elizabeth
Sent: Monday, September 13, 1999 5:22 PM
To: STIC-Biotech/ChemLib
Subject: 09/068,507

This is a date case, please let me know if I need to do anything to facilitate.

Please search for case **09/068,507**:

1. Seq ID NOs 1 + 2 against commercial ~~and interference~~ databases.

Thank you very much.

Elizabeth Slobodyansky

AM 1652
CM1-10D11
tel. 306-3222

P

IMPORTANT INFORMATION ABOUT YOUR SEQUENCE SEARCH:

Compugen Sequence searching hardware and software explained:

This is the new sequence searching system that is currently being phased into as a replacement for the Maspar/Mpsrch platform. This system has been tested by both searchers and examiners, and has shown equivalent results to the Maspar system for the same databases. The results output format for all Compugen printed results are essentially the same except for translations.

Translation searching on Compugen explained:

The Compugen system utilizes Framesearch software for translations of proteins to nucleotides, and nucleotides to proteins. Some examiners have found these to be superior to the backtranslate software on Maspars.

FrameSearch searches a group of protein sequences for similarity to one or more nucleotide query sequences, or searches a group of nucleotide sequences for similarity to one or more protein query sequences. For each sequence comparison, the program finds an optimal alignment between the protein sequence and the corresponding codons on each the nucleotide sequence. Optimal alignments may include reading frame shifts. Please see any of the professional searching staff if you need assistance with this format.

File extensions for Compugen results transferred to floppy disks.

Compugen system search results will be delivered in one of two possible formats:

1. Standard concatenated files with .flp extension.
2. Compressed .zip files which decompressed yield two files as described below:

US08123456.cmr - Contains all commercial databases, may include Issued
US08123456.pen - Contains pending file results only

VERY IMPORTANT NOTE ABOUT PENDING FILE SEARCHES.

If your search contains file names with the following bolded extensions:

US08123456.rap US08123456.rnp

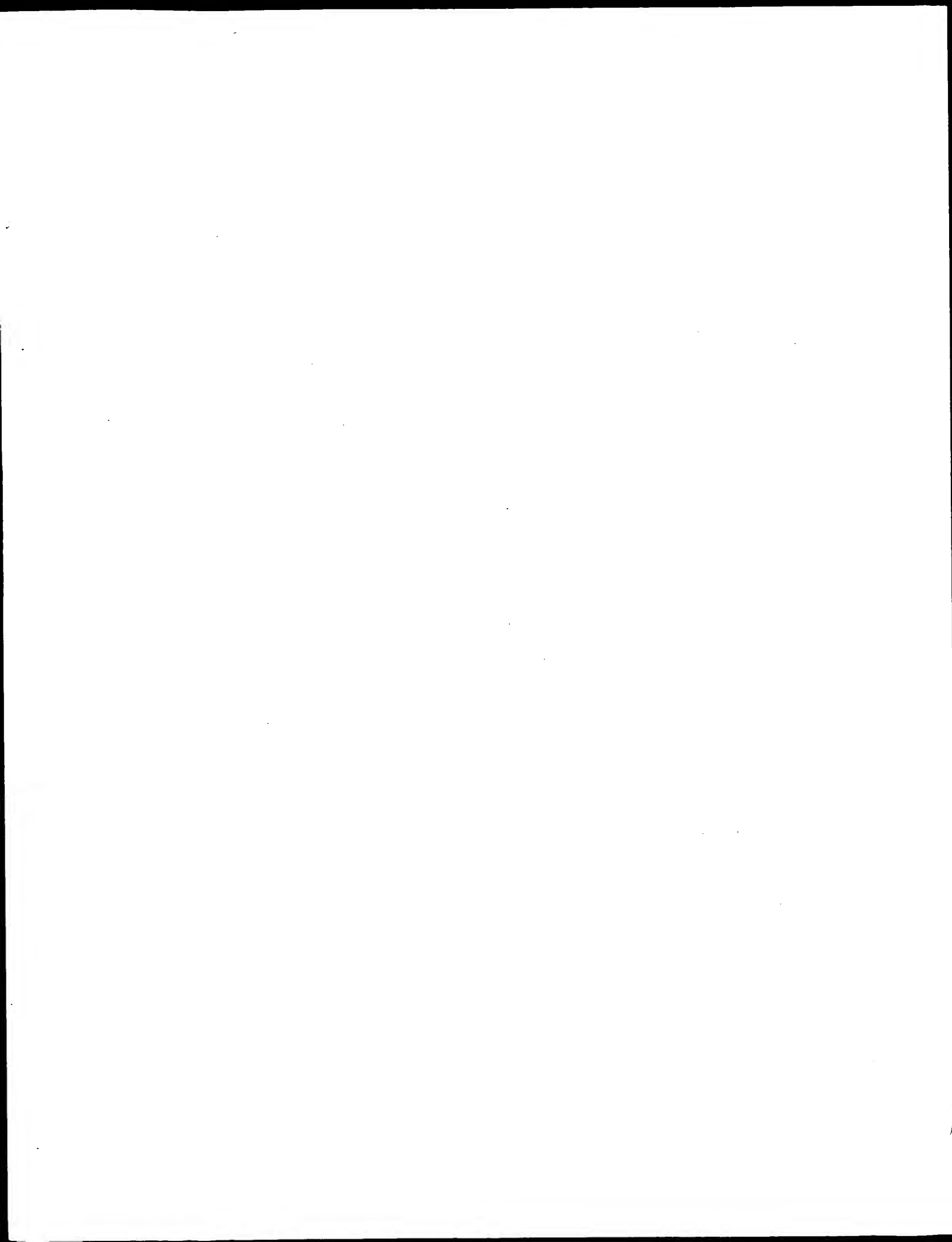
Do not leave this search in the case, during prosecution, or after the case issues, since it contains pending data which is confidential.

QUESTIONS? Contact any of the following:

Dilip Pandya, Chief, Information Branch, 308-4268

Professional searching staff:

John Dantzman (308-4488); Jan Delaval (308-4498); Mary Hale (308-4258); Barb O'Bryen (308-4291); David Schreiber (308-4292); Paula Sheppard (308-4499); Mark Spencer (308-4266); Beverly Shears (308-4994); Alex Wacławiw (308-4491).



FOR OFFICIAL USE ONLY

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office195-19
SEARCH REQUEST FORM

Examiner # (Mandatory): _____ Requester's Full Name: _____

Art Unit _____ Location (Bldg/Room#): 10011 Phone (circle 305 306 308) _____

Serial Number: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms): _____

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: BOBSearcher Phone #: 303-4740

Searcher Location: _____

Date Picked Up: 9-15

Date Completed: _____

Clerical Prep Time: _____

Terminal Time: _____

Number of Databases: _____

Type of Search

1 N.A. Sequence1 A.A. Sequence

____ Structure (#)

____ Bibliographic

____ Litigation1

____ Fulltext

____ Procurement

____ Other

Vendors (include cost where applicable)

____ STN

____ Questel/Orbit

____ Lexis/Nexis

____ WWW/Internet

AP 2 In-house sequence systems (list)

____ Dialog

____ Dr. Link

____ Westlaw

____ Other (specify)



Query Match 100.0%; Score 199; DB 24; Length 26;

Best Local Similarity 100.0%; Pred. No. 2, 87e-13;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 kssayslmgataikvkkkfkxgw 26

OY 1 KSSAYSLMGATAIKOVKKLFFKMGW 26

RESULT 2

ID W69847 standard; Protein: 822 AA.

AC W69847;

DT 07-OCT-1998 (first entry)

DE Amino acid sequence of p96, a protein present in SA-17S complex.

KM p96 protein: secretion associated 17S complex; SA-17S.

KM syntaxin-containing complex; SC complex; screening; modulate;

KM vesicular release; synaptic transmission; secretory process;

KM treatment; affective disorder; depression; manic-depressive disorder;

KM anxiety disorder; neurodegenerative disease; schizophrenia; anaesthesia;

KM hormonal imbalance; antigen processing; ss.

OS Rattus sp.

FT Key Location/Qualifiers

FT Misc.difference 412 /note="not specified"

PN W09828419-A2.

PD 02-JUL-1998.

PF 19-DEC-1997; U23498.

PR 20-DEC-1996; US-033905.

PA (STRD) UNIV LELAND STANFORD JUNIOR.

PI Scheller RH;

DR WPI: 98-377650/32.

DR N-PSDB: V42652.

PT New isolated vesicle secretion associated poly(peptide(s) - used to

PT develop products for treating e.g. effective disorder,

PT neurodegenerative disease, hormone imbalances, immune system

PT disorders or tumours

PS Claim 1d: Pages 107-109; 133pp; English.

CC The present sequence represents a p96 protein. This protein is present

CC SA-17S complex, which binds a syntaxin-containing (SC) complex. The

CC SA-17S polypeptides and nucleotide sequences encoding them can be used

CC for screening for compounds which modulate vesicular release involved in

CC synaptic transmission and other secretory processes. Compounds which

CC enhance binding between the SA-17S and SC complexes may be used to treat

CC an affective disorder such as depression, manic-depressive disorders and

CC anxiety disorders, or a neurodegenerative disease such as Parkinson's

CC disease or Huntington's disease. Compounds which inhibit binding between

CC the SA-17S and SC complexes may be used to treat a disorder of thought,

CC such as schizophrenia, or for anaesthesia. The compounds can also be used

CC to intervene in the endocrine system for treatment of hormonal

CC imbalances, the immune system for intervention in antigen processing,

CC secreted immunomodulators, and viral processing, as well as

CC anti-tumour applications, such as regulation of membrane trafficking

CC during rapid cell division.

CC Sequence 822 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1b. GABA (gamma-aminobutyric acid) is the major inhibitory

CC neurotransmitter found in the brain and peripheral nervous system

CC and this receptor may be used for the identification of GABA-B

CC receptor agonists and antagonists. Such proteins may be used in

CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.

CC GABA-B receptor ligands and probes derived from this sequence can be

CC used to assay for GABA-B receptors or DNA encoding them.

CC Sequence 793 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1b. GABA (gamma-aminobutyric acid) is the major inhibitory

CC neurotransmitter found in the brain and peripheral nervous system

CC and this receptor may be used for the identification of GABA-B

CC receptor agonists and antagonists. Such proteins may be used in

CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.

CC GABA-B receptor ligands and probes derived from this sequence can be

CC used to assay for GABA-B receptors or DNA encoding them.

CC Sequence 793 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1b. GABA (gamma-aminobutyric acid) is the major inhibitory

CC neurotransmitter found in the brain and peripheral nervous system

CC and this receptor may be used for the identification of GABA-B

CC receptor agonists and antagonists. Such proteins may be used in

CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.

CC GABA-B receptor ligands and probes derived from this sequence can be

CC used to assay for GABA-B receptors or DNA encoding them.

CC Sequence 793 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1b. GABA (gamma-aminobutyric acid) is the major inhibitory

OY 19 KLFKKMGW 26

RESULT 5
ID W40119 standard; Protein: 844 AA.

AC W40119;
DT 03-JUN-1998 (first entry)
DE Human GABA-BR1b receptor protein.
KW Gamma-aminobutyric acid; GABA-BR1a/b receptor; human; brain; agonist;
KW Inhibitory neurotransmitter; peripheral nervous system; antagonist;
KW treatment; dementia; depression; anxiety; bronchial inflammation; asthma;
KW epilepsy; cognitive function.
OS Homo sapiens.
PN W09746675-A1.
PD 11-DEC-1997.
PE 19-MAR-1997; E01370.
PR 22-NOV-1996; US-756091.
PR 30-MAY-1996; US-655716.
PA (NOVS) NOVARTIS AG.
PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;
WPI: 98-042183/04.
DR N-PSDB: V10267.
PT Purified GABA-B receptor or receptor protein - and antagonists of
these which may be useful in treating nervous system disorders

PS Claim 4; Page 86-90; 108pp; English.
CC This sequence represents a novel human GABA-B receptor protein,
GABA-BR1b. GABA (gamma-aminobutyric acid) is the major inhibitory
neurotransmitter found in the brain and peripheral nervous system
and this receptor may be used for the identification of GABA-B
receptor agonists and antagonists. Such proteins may be used in
treatment of dementia, depression, anxiety, epilepsy, spasticity,
bronchial inflammation or asthma or to improve cognitive function.
CC GABA-B receptor ligands and probes derived from this sequence can be
used to assay for GABA-B receptors or DNA encoding them.
SQ Sequence 844 AA;

Query Match 37.7%; Score 75; DB 29; Length 844;
Best Local Similarity 87.5%; Pred. No. 6.84e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 182 klfekgw 189

OY 19 KLFKKMGW 26

RESULT 6
ID W40116 standard; Protein: 960 AA.

AC W40116;
DT 03-JUN-1998 (first entry)
DE Rat GABA-BR1a receptor protein.
KW Gamma-aminobutyric acid; GABA-BR1a receptor; rat; brain; agonist;
KW Inhibitory neurotransmitter; peripheral nervous system; antagonist;
KW treatment; dementia; depression; anxiety; bronchial inflammation; asthma;
KW epilepsy; cognitive function.
OS Rattus norvegicus.
PN W09746675-A1.
PD 11-DEC-1997.
PE 19-MAR-1997; E01370.
PR 22-NOV-1996; US-756091.
PR 30-MAY-1996; US-655716.
PA (NOVS) NOVARTIS AG.
PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;
WPI: 98-042183/04.
DR N-PSDB: V10264.
PT Purified GABA-B receptor or receptor protein - and antagonists of
these which may be useful in treating nervous system disorders

PS Claim 4; Page 50-56; 108pp; English.
CC This sequence represents a novel rat GABA-B receptor protein,
GABA-BR1a. GABA (gamma-aminobutyric acid) is the major inhibitory
neurotransmitter found in the brain and peripheral nervous system
and this receptor may be used for the identification of GABA-B
receptor agonists and antagonists. Such proteins may be used in
treatment of dementia, depression, anxiety, epilepsy, spasticity,
bronchial inflammation or asthma or to improve cognitive function.

CC bronchial inflammation or asthma or to improve cognitive function.
CC GABA-B receptor ligands and probes derived from this sequence can be
used to assay for GABA-B receptors or DNA encoding them.
SQ Sequence 960 AA.

Query Match 37.7%; Score 75; DB 29; Length 960;
Best Local Similarity 87.5%; Pred. No. 6.84e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 298 klfekgw 305

OY 19 KLFKKMGW 26

RESULT 7
ID R78520 standard; Protein: 572 AA.

AC R78520;
DT 15-NOV-1995 (first entry)
DE Partial ALK protein.
KW ALK protein; protein-tyrosine-kinase; fusion protein; NPM protein;
KW nucleolar phosphoprotein; anaplastic large cell lymphoma;
KW t(2;5) translocation; diagnosis; gene therapy.
OS Homo sapiens.
PN W0951331-A.
PD 08-JUN-1995.
PE 05-DEC-1994; U13947.
PR 03-DEC-1993; US-160861.
PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PI Look AT, Morris SW;
WPI: 95-215226/28.
DR N-PSDB: Q95341.
PT Methods for detecting human t(2;5) lymphoma - for detection and
diagnosis of anaplastic large cell lymphoma(s)
PS Claim 21; Page 40-42; 70pp; English.
CC The translocation event that occurs in human t(2;5) lymphoma
brings sequences from the nucleolar phosphoprotein (NPM) gene on
chromosome 3q35 to those from a protein-tyrosine-kinase (ALK) gene
on chromosome-2q23. Nucleic acids encoding a partial ALK protein
(R78520) and the ALK/NPM fusion protein (R78521) were isolated
(Q95541, Q95542). Identification of the NPM/ALK fusion allows
CC lymphoma diagnosis, or therapy through the use of antisense RNA or
ribozymes.
SQ Sequence 572 AA;

Query Match 35.7%; Score 71; DB 13; Length 572;
Best Local Similarity 52.4%; Pred. No. 1.64e+01;
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 2 slqegatgshcpgamkkgw 22

OY 6 SLQMGATAIKQVKLKKMGW 26

RESULT 8
ID W62742 standard; Protein: 234 AA.

AC W62742;
DT 09-NOV-1998 (first entry)
DE Streptococcus pneumoniae polypeptide.
KW Polypeptide; ORF; open reading frame; infection; bacterial;
KW streptococcal; bacteremia; diagnosis; prophylaxis.
OS Streptococcus pneumoniae.
PN W09823651-A1.
PD 04-JUN-1998.
PE 24-NOV-1997; U21976.
PR 27-NOV-1996; US-031879.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Black MR, Hodgson JE, Knowles DJC, Ionetto MA, Nicholas RO,
WPI: 98-322654/28.
DR Streptococcus pneumoniae polynucleotides - useful for developing
PT products for diagnosis, prevention and treatment of infections e.g.
pneumonia, bacteremia, meningitis or endocarditis

RESULT 10
ID W41077 standard; Protein; 467 AA.
C W41077;
P 05-MAY-1998 (first entry)
DE 7. Iltoralis ADPKK protein sequence, SEQ ID NO 2.
E ADPKK; enzyme; hexokinase; hexose 6'-phosphate; adenosine 1'-phosphate;
W adenosine 2'-phosphate; hexose.
S Thermococcus Iltoralis.
J09337297-A.

RESULT	12
ID	R91047 standard; Protein; 1577 AA.
AC	R91047.
DT	22-MAY-1996 (first entry)
DR	Alpha-D-glucosyltransferase.
EC	Alpha-D-glucosyltransferase.
KW	sucrose; transgenic plant; cloning; Escherichia coli;
KW	phage lambda C13; vector; plasmid pSGG502;
KW	gene transfer; crop improvement; storage carbohydrate; pasture;

RESULT 13
ID W12897 standard: peptide; 13 AA.
AC W12897:
DT 10-DEC-1997 (first entry)
DE Antimicrobial cationic peptide Cp-23.
KW Bacterial; viral; antitumour; food; preservative; inhibitor; growth;
bacterium; yeast; endotoxaemia; sepsis; antibiotic; fungal;
antiviral; Candida albicans; sterlant; Salmonella; Yersinia;
Shigella.
OS Synthetic.
PN W09708199-A2.
PD 06-MAR-1997.
PF 23-AUG-1996; IB0996.
PR 23-AUG-1995; US-002687.
PA (UYBR-1) UNIV BRITISH COLUMBIA.
PI Falla TJ, Gough M, Hancock REM:
WPI: 97-179179/16.
PT Cationic peptide(s) having anti-microbial activity - used for the
PT inhibition of bacterial and viral growth, as an antitumour agent,
PT and as a food preservative
PS Claim 3; Page 66; 89pp; English.
CC The present sequence represents a specifically claimed novel isolated
CC cationic peptide which has antimicrobial activity. The amino acid
CC sequence of antimicrobial cationic peptide(s) including the present
CC sequence is selected from: X1X1Prox2XxX2Pro(x2XxPro)nX2X3(X5)O;
X1X1Prox2X3X4(X5)irProx2X3X3; X1X1X3(PriOTir)uX3X2X5X2X5X2(X5)O;
X1X1X3X3X3Pir(x2X2Pro)nX2(X5)m; where m = 1-5; n = 1-2; o = 2-5; i
CC = 0-8; u = 0-1; X1 = Ile, Leu, Val, Phe, Tyr, Trp or Met; X2 = Trp or
CC Phe; X3 = Arg or Lys; X4 = Trp or Lys; and X5 = Phe, Trp, Arg, Lys or
CC Pro. The peptides are preferably amidated or carboxymethylated. The
CC peptides may be used in methods for inhibiting the growth of a bacterium
CC or yeast, or for inhibiting an endotoxaemia or sepsis associated
CC disorder in a subject. The peptides have a broad activity against

RESULT	15
ID	W61021 standard; Protein, 128 AA.
AC	W61021;
DR	13-OCT-1998 (first entry)
DE	<i>Streptococcus pneumoniae</i> encoded polypeptide

KW coding region: ORF: open reading frame: antibacterial;
 KM infection: prevention: meningitis.
 OS Streptococcus pneumoniae.
 PN W09819689-A1.
 PD 14-MAY-1998.
 PF 27-OCT-1997; 019226.
 PR 01-NOV-1996; US-029930.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black Mt, Hodgson JE, Knowles DUC, Lonetto MA, Nicholas RO,
 PI Reid RH, Zarfos PN;
 DR WPI: 98-286586/25.
 DR N-PSDB; V37409.
 PT New isolated nucleic acids from Streptococcus pneumoniae - useful,
 PT e.g. for identifying anti-bacterial(s) for treatment and prevention
 PT of meningitis
 PS Claim 11; Page 114; 130pp; English.
 CC The sequence is that of the polypeptide encoded by a region isolated
 CC from S. pneumoniae. The protein, or agonists of it,
 CC may be useful as an antibacterial for treatment or
 CC prevention of infection, specifically caused by S. pneumoniae
 CC (particularly meningitis) but possibly also Helicobacter
 CC pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other
 CC invasive procedure. The protein, or nucleic acid encoding
 CC it, can also be used in vaccines to induce a cellular
 CC and/or humoral immune response, or to screen for other
 CC antibacterials. The DNA may also contain flanking sequences
 CC that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein
 CC can be used diagnostically, e.g. to detect a mutation for
 CC serotyping or classifying infectious agents.
 SQ Sequence 128 AA:

Query Match 30.28; Score 60; DB 33; Length 128;
 Best Local Similarity 53.88; Pred. No. 1.67e+02;
 Matches 7; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Db. 60 kkaerdlfkfw-w 71
 | | : | | | | |
 Oy 14 IKOVKKLKKRW 26

Search completed: Wed Sep 15 10:38:07 1999
 Job time : 17 secs.

W A S E R H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:40:03 1999; MasPar time 1.59 seconds
Tabular output not generated. 165.973 Million cell updates/sec

Title: >US-09-068-507A-1
Description: (1-26) from US09068507A.pep
Perfect Score: 199
Sequence: 1 KSSAYSLQMGATAIKQVKLFKKWG 26

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 21.296; Variance 76.695; scale 0.278

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	71	35.7	572	1 US-08-160-	Sequence 3, Applicatio	8.35e+00
2	71	35.7	1620	2 US-08-542-	Sequence 2, Applicatio	8.35e+00
3	60	30.2	572	2 US-08-185-	Sequence 81, Applicati	8.17e+01
4	60	30.2	572	2 US-08-191-	Sequence 81, Applicati	8.17e+01
5	59	29.6	231	2 US-08-808-	Sequence 33, Applicati	9.99e+01
6	59	29.6	316	3 PCT-US95-0	Sequence 2, Applicatio	9.99e+01
7	59	29.6	316	2 US-08-464-	Sequence 2, Applicatio	9.99e+01
8	59	29.6	372	4 5514582-2	Patent No. 5514582.	9.99e+01
9	59	29.6	372	2 US-08-513-	Sequence 2, Applicatio	9.99e+01
10	59	29.6	385	2 US-08-340-	Sequence 2, Applicatio	9.99e+01
11	59	29.6	385	2 US-08-461-	Sequence 2, Applicatio	9.99e+01
12	58	29.1	410	1 US-08-073-	Sequence 16, Applicati	1.12e+02
13	57	28.6	49	2 US-08-456-	Sequence 26, Applicati	1.49e+02
14	57	28.6	49	2 US-08-237-	Sequence 26, Applicati	1.49e+02
15	57	28.6	450	3 PCT-US95-0	Sequence 7, Applicati	1.49e+02
16	57	28.6	485	1 US-07-991-	Sequence 42, Applicati	1.49e+02
17	57	28.6	544	1 US-08-264-	Sequence 7, Applicatio	1.49e+02
18	57	28.6	1167	1 US-08-485-	Sequence 6, Applicatio	1.49e+02
19	57	28.6	1167	1 US-08-590-	Sequence 6, Applicatio	1.49e+02
20	57	28.6	1168	1 US-08-620-	Sequence 9, Applicatio	1.49e+02
21	57	28.6	1220	2 US-08-158-	Sequence 43, Applicati	1.49e+02
22	57	28.6	1220	2 US-08-611-	Sequence 43, Applicati	1.49e+02
23	57	28.6	1289	4 5281530-3	Patent No. 5281530.	1.49e+02

24	57	28.6	1289	4 5426049-4	Patent No. 5426049.	1.49e+02
25	57	28.6	1289	1 US-08-158-	Sequence 4, Applicatio	1.49e+02
26	57	28.6	1289	1 US-07-876-	Sequence 4, Applicatio	1.49e+02
27	57	28.6	1289	1 US-08-083-	Sequence 4, Applicatio	1.49e+02
28	57	28.6	1289	2 US-08-316-	Sequence 4, Applicatio	1.49e+02
29	57	28.6	1289	1 US-08-304-	Sequence 4, Applicatio	1.49e+02
30	57	28.6	1289	3 PCT-US92-0	Sequence 4, Applicatio	1.49e+02
31	57	28.6	1289	1 US-07-675-	Sequence 4, Applicatio	1.49e+02
32	57	28.6	1289	2 US-08-611-	Sequence 4, Applicatio	1.49e+02
33	57	28.6	1385	4 5426049-1	Patent No. 5426049.	1.49e+02
34	57	28.6	1385	1 US-07-876-	Sequence 2, Applicatio	1.49e+02
35	57	28.6	1385	2 US-08-316-	Sequence 2, Applicatio	1.49e+02
36	57	28.6	1385	4 5281530-1	Patent No. 5281530.	1.49e+02
37	57	28.6	1385	1 US-07-675-	Sequence 2, Applicatio	1.49e+02
38	57	28.6	1385	2 US-08-611-	Sequence 2, Applicatio	1.49e+02
39	57	28.6	1385	1 US-08-083-	Sequence 2, Applicatio	1.49e+02
40	57	28.6	1385	1 US-08-158-	Sequence 2, Applicatio	1.49e+02
41	57	28.6	1385	3 PCT-US92-0	Sequence 2, Applicatio	1.49e+02
42	57	28.6	1385	1 US-08-304-	Sequence 2, Applicatio	1.49e+02
43	56	28.1	119	2 US-08-340-	Sequence 19, Applicati	1.81e+02
44	56	28.1	338	1 US-08-553-	Sequence 16, Applicati	1.81e+02
45	56	28.1	1023	1 US-08-198-	Sequence 4, Applicatio	1.81e+02

ALIGNMENTS

RESULT 1
ID US-08-160-861-3 STANDARD: PRT: 572 AA.

XX
AC xxxxxx

XX
DT

XX
DE

XX
XX

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

CC
CC

Sequence 3, Application US/08160861

Sequence 3, Application US/08160861

Patent No. 5529295

GENERAL INFORMATION:

APPLICANT: MORRIS, STEPHAN W

TITLE OF INVENTION: NOVEL FUSION NUCLEIC ACID SEQUENCES AND

TITLE OF INVENTION: FUSION PROTEINS PRESENT IN HUMAN t(2:5) LYMPHOMA, METHO

TITLE OF INVENTION: OF DETECTION AND USES THEREOF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX

STREET: 1100 NEW YORK AVE NW SUITE 600

CITY: WASHINGTON

STATE: D.C.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/160.861

FILING DATE: 02-DEC-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MILLMAN, ROBERT A

REGISTRATION NUMBER: 36217

REFERENCE/DOCKET NUMBER: 0656.0400000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2678

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 572 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 572 AA: 62849 MW; 1722781 CN;

Query Match 35.7%; Score 71; DB 1; Length 572;

Best Local Similarity 52.4%; Pred. No. 8.35e+00;
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 2 SLOEGATGHHSCPOAMKKWGW 22
||| |||
QY 6 SLOMGATAIKOVKKLFKKWGW 26

RESULT 2
ID US-08-542-363-2 STANDARD; PRT; 1620 AA.

XX

AC

XX

DT

XX

DE Sequence 2, Application US/08542363

XX Sequence 2, Application US/08542363

CC Patent No. 5770421

CC GENERAL INFORMATION:

CC APPLICANT: Morris, Stephan W.

CC APPLICANT: Look, A. Thomas

CC TITLE OF INVENTION: ALK Protein Tyrosine Kinase/Receptor and

CC TITLE OF INVENTION: Ligands Thereof

CC NUMBER OF SEQUENCES: 43

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

CC STREET: 1100 New York Avenue, N.W., Suite 600

CC CITY: Washington

CC STATE: DC

CC COUNTRY: USA

CC ZIP: 20005

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: IBM PC compatible

CC SOFTWARE: PC-DOS/MS-DOS

CC OPERATING SYSTEM: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/542,363

CC FILING DATE: 12-OCT-1995

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fox, Samuel L.

CC REGISTRATION NUMBER: 30,353

CC REFERENCE/DOCKET NUMBER: 0656.0400001/SLF/GKT

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 202-371-2600

CC TELEFAX: 202-371-2540

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1620 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 1620 AA; 176416 MW; 13627521 CN;

Query Match 35.7%; Score 71; DB 2; Length 1620;

Best Local Similarity 52.4%; Pred. No. 8.35e+00;

Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 895 SLOEGATGHHSCPOAMKKWGW 915

||| |||

QY 6 SLOMGATAIKOVKKLFKKWGW 26

RESULT 3

ID US-08-185-949B-81 STANDARD; PRT; 572 AA.

XX

AC

XX

DT

XX

DE Sequence 81, Application US/08185949B

XX Sequence 81, Application US/08185949B

CC Patent No. 5874279

CC GENERAL INFORMATION:

CC APPLICANT: Mark D. Cochran

CC APPLICANT: Richard D. Macdonald

CC TITLE OF INVENTION: Recombinant Infectious Bovine

CC TITLE OF INVENTION: Rhinotracheitis Virus

CC NUMBER OF SEQUENCES: 104

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: John P. White

CC STREET: 1185 Avenue of the Americas

CC CITY: New York

CC STATE: New York

CC COUNTRY: USA

CC ZIP: 10036

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM 330 466 DX2

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/185,949B

CC FILING DATE: 03-NOV-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: White, John P.

CC REGISTRATION NUMBER: .678

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 278-0400

CC TELEFAX: (212) 278-0525

CC INFORMATION FOR SEQ ID NO: 81:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 572 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 572 AA; 64607 MW; 1762281 CN;

Query Match 30.2%; Score 60; DB 2; Length 572;

Best Local Similarity 30.0%; Pred. No. 8.17e+01;

Matches 6; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 432 LQLGVIDISYNNIRINWTW 451

||| | : : : |

QY 7 LQMGATAIKOVKKLFKKWGW 26

RESULT 4

ID US-08-191-866D-81 STANDARD; PRT; 572 AA.

XX

AC

XX

DT

XX

DE Sequence 81, Application US/08191866D

XX Sequence 81, Application US/08191866D

CC Patent No. 5783195

CC GENERAL INFORMATION:

CC APPLICANT: Cochran, Mark D

CC APPLICANT: Macdonald, Richard D.

CC TITLE OF INVENTION: Recombinant Infectious Bovine

CC TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof

CC NUMBER OF SEQUENCES: 99

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: John P. White

CC STREET: 1185 Avenue of the Americas

CC CITY: New York

CC STATE: New York

CC COUNTRY: USA

CC ZIP: 10036

CN;
 9; DB 2; Length 231;
 C. 9,99e+01;
 smatches 3; Indels 0; Gaps 0;

 PRT; 316 AA.

 27A
 827A
 atatic Specific Reductase
 HAIN, GILFILLAN,
 ; OLSTEIN

 TTE

 5/01827A

 5800-228

 4 CN;


```
CC CC  
CC CC  
CC CC  
CC CC  
CC CC  
CC CC  
SQ
```

SEQUENCE CHARACTERISTICS:
LENGTH: 385 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 385 AA; 43743 MW; 775024 CN;

Query Match 29.6%; Score 59; DB 2: Length 395;
Best Local Similarity 62.5%; Pred. No. 9,99e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC APPLICATION NUMBER: 07/786149
CC FILING DATE: 31-OCT-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/315015
CC FILING DATE: 23-FEB-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Dreger, Ginger R.
CC REGISTRATION NUMBER: 33,055

```
db      29 NIEKLGW 36
        :|| |||
QY      19 KLFKKGW 26
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881

RESULT	11	PRT;	385 AA.
		STANDARD:	

INFORMATION ON SEQUENCE CHARACTERISTICS:

AA
AC
xxxxxxx

TYPE: linear
TOPOLOGY: linear
T00050 CN.

DT
XX
XX 00461503B

29.6%; score 59; DB 2; Length 3/2,

XX
CC
sequence 2, Application US/08461592B

Matches

CC GENERAL INFORMATION: Thomas F.
CC APPLICANT: Tedder, Thomas F.

20

CC TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTION FOR

CC
CC
CC
CORRESPONDENCE ADDRESS:
CORRESPONDENCE: Weingarten, Schurgin, Gagnebin & Hayes

28

CITY: Boston

2000

ZIP: 02109

DE sequence 2, nppzcc

COMPUTER: IBM PC compatible
OS: DOS/MS-DOS

CC Patent No. 3808025
CC GENERAL INFORMATION:

SOFIWARE.
CURRENT APPLICATION DATA:
502B

CC

APPLICANT: Kansas, Secretary of State
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS COMPONENT SELECTIN FUNCTION

CC
CC
CC
FILING DATE: 514
CLASSIFICATION: 514

CC
NUMBER OF SEQUENCES: 20
SEQUENCE ADDRESS:

CC
APPLICATION NUMBER: CC 89-
FILING DATE: 16-NOV-1994

STREET: 1251 Avenue of the Americas
New York

APPLICATION NUMBER: 03 007 0007
 RECEIVING DATE: 25-JAN-1993

COUNTRY: USA

NAME: James F. Haley, Jr.
STORY NUMBER: 27:794

CC
COMPUTER NAME: FLOPPY disk
MEDIUM TYPE: FLOPPY disk
CC

TELECOMMUNICATION INFORMATION:
505-8000

CC
CC
CC
OPERATING SYSTEM: PatentIn Release #1.0, Version #1.30
SOFTWARE:

TELEFAX: 14-8367
TELEX: 14-8367

CC
APPLICATION NUMBER: 16-NOV-1994
FILING DATE: 16-NOV-1994

CC
CC
SEQUENCE CHARGE
LENGTH: 385 amino acids

PRIOR APPLICATION DATA: US 08/008,459

CC
 TOPOLOGY: linear
 MOLECULE TYPE: protein

ATTORNEY/AGENT INFORMATION:

29.6%: Score 59; DB 2; Length 385;

REFERENCE/DOCKET NUMBER: CG-104 CON
INFORMATION:

TELEPHONE: 212-596-9090
TELEFAX: 212-596-9090

1

	Matches	5; Conservative	2; Mismatches	1; Indels	0; Gaps	0;
Db	29	NIFKLWGW	36			
		: : : : :				
Qy	19	KLFFKKGW	26			

RESULT 12
ID US-08-073-807A-16 STANDARD; PRT; 410 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 16, APPLICATION US/08073807A
CC SEQUENCE 16, APPLICATION US/08073807A
CC PATENT NO. 5646248
CC GENERAL INFORMATION:
CC APPLICANT: Sawada, Ritsuko
CC APPLICANT: Lowe, John B.
CC APPLICANT: Fukuda, Minoru
CC TITLE OF INVENTION: CELL SURFACE LAMP EXPRESSION AND
CC TITLE OF INVENTION: SELECTIN-DEPENDENT ADHESION
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC

ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,807A
FILING DATE: 08-JUN-1993
CLASS: 08

ATTORNEY/AGENT INFORMATION: 435
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9567
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear

SEQUENCE 410 AA; 4988 MW; 908539 CN;

Query Match 29.1% Score 58; DB 1; Length 410;
Best Local Similarity 28.6% Pred. NO. 1.22e-02;
Matches 6; Conservative 5; Mismatches 10. 7-2-21-

Db	26	SVAFLENI	TCCTGTTATTT	Matches	10;	Indels	0;	Gaps	0;
----	----	----------	-------------	---------	-----	--------	----	------	----

QY
4 AYSLQMGATAIKQVKKLFKKW 24

RESULT 13
ID US-08-456-647B-26
XX
AC xxxxxx
STANDARD; PRT; 49 AA.

Sequence 26, Application US/08456647B

1

Sequence 26, Application US/08456647B
Patent No. 5811516
GENERAL INFORMATION:
APPLICANT: Lemke Ph.D. et al., Greg E.
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037

```
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/456,647B
 FILING DATE: 02-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/237,401
 FILING DATE: 02-MAY-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/884,486
 FILING DATE: 13-MAY-1992
 ATTORNEY/AGENT:

CC NAME: Wetherell Ph.D., John R.
CC REGISTRATION NUMBER: 31,678
CC REFERENCE/DOCKET NUMBER: 07251/007002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 678-5070
CC TELEFAX: (619) 678-5099
CC INFORMATION FOR SEQ ID NO: 26:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 49 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 49 AA, 5339 MW, 14108 CN;

Query Match	28.6%	Score 57;	DB 2;	Length 49;
Best Local Similarity	33.3%	Pred. NO. 1.49e+02;		
Matches	8;	Conservative	6;	Mismatches 10; Indels

Seq	Matches	Indels	Gaps
Db	11	0	0

QY 1 KSSAYSLOMGATAIKOVKKLEFKKW 24

RESULT	14	
ID	US-08-237-401A-26	
XX		STANDARD;
AC	XXXXXX	PRT; 49 AA.

Sequence 26, Application US/08237401A
Sequence 26, Application US/08237401A
Patent No. 5837448
GENERAL INFORMATION:
APPLICANT: Lemke Ph.D. et al., Greg E.
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US

```

CC ZIP: 92037
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/237,401A
CC FILING DATE: 02-MAY-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/884,486
CC FILING DATE: 15-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Haile Ph.D., Lisa A.
CC REGISTRATION NUMBER: 38,347
CC REFERENCE/DOCKET NUMBER: 07251/007001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 678-5070
CC TELEFAX: (619) 678-5099
CC INFORMATION FOR SEQ ID NO: 26:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 49 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 49 AA: 5339 MW: 14108 CN:

Query Match 28.68; Score 57; DB 2; Length
Best Local Similarity 33.38; Pred. No. 1.49e+02;
Matches 8; Conservative 6; Mismatches 10; In

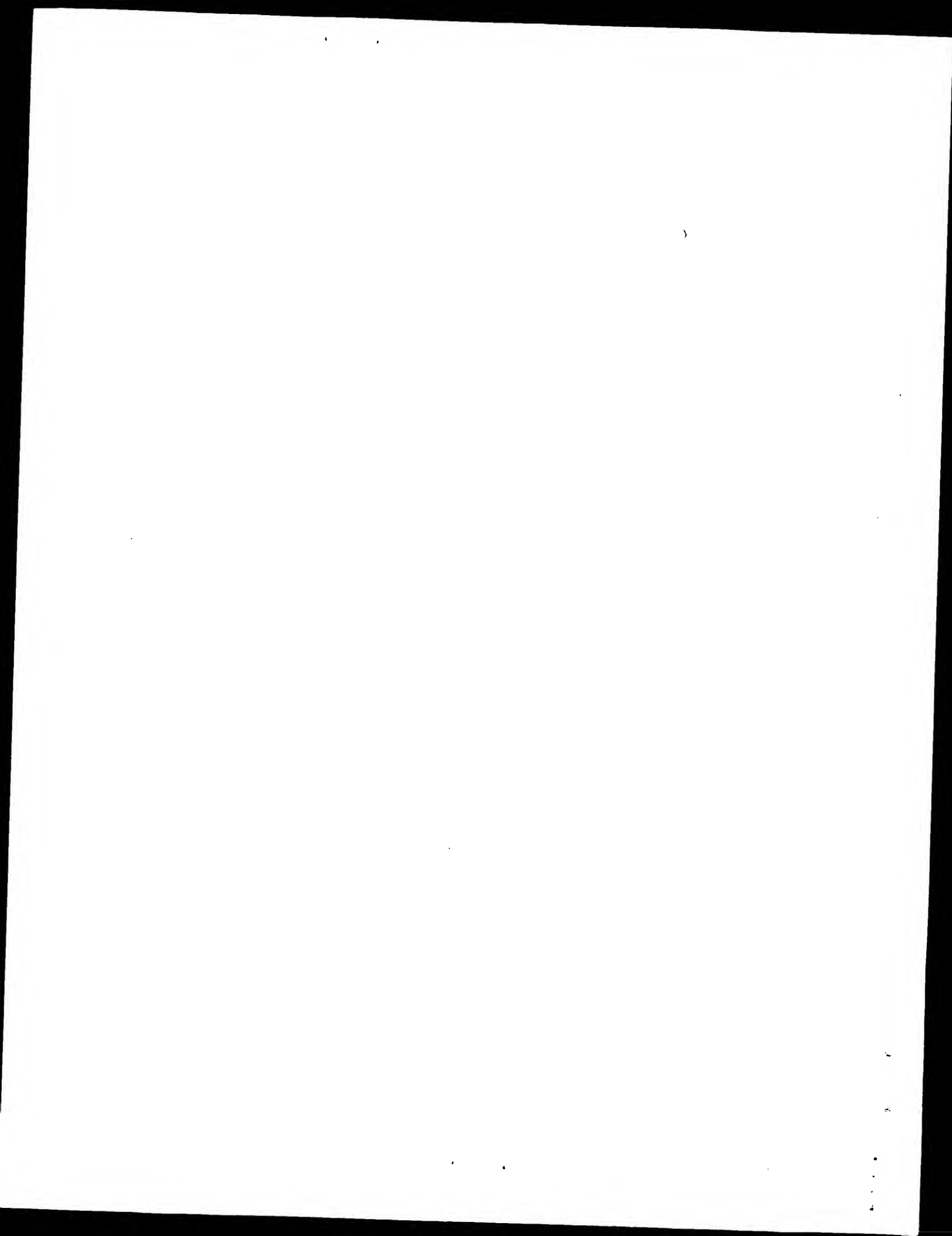
Db 11 KVSDFGLTKAESQTDTGKLPVKW 34
QY 1 KSSAYSLOWGATAIKQVKLPFKW 24
| | : : | : : | : |
| | : : | : : | : |

RESULT 15 STANDARD: PRT; 450 AA.
ID PCT-US95-05008-7
XX XXXXXX
AC
XX
XX
DT
XX
DE
XX
XX
CC Sequence 7, Application PC/TUS9505008
CC Sequence 7, Application PC/TUS9505008
CC GENERAL INFORMATION:
CC APPLICANT: Sugen, Inc.
CC APPLICANT: 515 Galveston Drive
CC APPLICANT: Redwood City, California 94063-4720
CC APPLICANT: United States of America
CC APPLICANT: Wissenschaften E.V.
CC APPLICANT: Hofgarten Str. 2
CC APPLICANT: Munchen 80539
CC APPLICANT: Germany
CC TITLE OF INVENTION: Novel Megakaryocytic Prote
CC NUMBER OF SEQUENCES: 21
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/05008

```

CC FILING DATE: 24-APR-1995
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/232,545
 CC FILING DATE: 22-APR-1994
 CC CLASSIFICATION:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Coruzzi, Laura A.
 CC REGISTRATION NUMBER: 30,742
 CC REFERENCE/DOCKET NUMBER: 7683-074
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (212)790-9090
 CC TELEFAX: (212)869-9741
 CC TELEX: 56141 PENNIE
 CC INFORMATION FOR SEQ ID NO: 7:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 450 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: unknown
 CC TOPOLOGY: unknown
 CC MOLECULE TYPE: protein
 CC SEQUENCE 450 AA; 50704 MW; 1071733 CN;
 SQ
 Query Match 28.6%; Score 57; DB 3: Length 450;
 Best Local Similarity 33.3%; Pred. No. 1.49e+02;
 Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0
 Db 329 KYSDFLTKKASSTQDTGKLPVKW 352
 |||::||::||::||
 QY 1 KSSAYSLQMGATAIKQVKLFFKW 24

Search completed: Wed Sep 15 10:40:11 1999
 Job time : 8 secs.



 W P S R L H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Sep 15 10:38:24 1999; MasPar time 4.14 Seconds
 Tabular output not generated. 251.685 Million cell updates/sec

Title: >US-09-068-507A-1
 Description: (1-26) from US09068507A.pap
 Perfect Score: 199
 Sequence: 1 KSSAYSQMGATAIKOVKKLFKKWG 26
 Scoring table: PAM 150
 Gap 15
 Searched: 122810 seqs, 40068593 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Database: pir60
 1:pir1 2:pir2 3:pir3 4:pir4
 Statistics: Mean 31.155; Variance 49.460; scale 0.630

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	166	83.4	22	2	A45913	plantaricin A - Lacto	5.88e-20
2	78	39.2	1609	2	S25345	probable membrane pro	7.06e-02
3	76	38.2	655	2	A42420	L-iduronidase (EC 3.2	1.56e-01
4	75	37.7	650	2	D71021	hypothetical protein	2.30e-01
5	75	37.7	960	2	JF0356	gamma-aminobutyric ac	2.30e-01
6	70	35.2	32	2	A40361	virc-region hypotherm	1.55e+00
7	69	34.7	649	2	B38129	bo-type ubiquinol oxi	2.25e+00
8	68	34.2	134	2	B2727	gedA protein - Mycoba	3.25e+00
9	68	34.2	157	2	C71050	hypothetical protein	3.25e+00
10	68	34.2	256	2	S41185	gene 38 protein - pha	4.68e+00
11	67	33.7	341	2	S31236	hypothetical protein	4.68e+00
12	67	33.7	457	2	A71174	hypothetical protein	6.70e+00
13	66	33.2	246	2	F69771	conserved hypotherm	6.70e+00
14	66	33.2	435	2	I52374	cerebrin-50 - human	6.70e+00
15	66	33.2	996	2	S42208	NAD+ ADP-ribosyltrans	9.58e+00
16	65	32.7	400	2	A27708	creatine kinase (EC 2	1.36e+01
17	64	32.2	529	2	I39841	transcription regulat	1.36e+01
18	64	32.2	835	2	F71638	hypothetical protein	1.93e+01
19	63	31.7	265	2	C70221	conserved hypotherm	1.93e+01
20	63	31.7	340	2	PH0217	reverse transcriptase	1.93e+01
21	63	31.7	366	2	C69391	enolase (eno) homolog	1.93e+01
22	63	31.7	419	2	S43989	gene 148 protein - fr	1.93e+01
23	63	31.7	486	2	JC4240	lanosterol 14alpha-de	1.93e+01

24	63	31.7	513	2	S21976	probable RNA-directed	1.93e+01
25	63	31.7	556	2	S21347	hypothetical protein	1.93e+01
26	63	31.7	685	2	S16783	probable RNA-directed	1.93e+01
27	63	31.7	809	2	S40460	ribosomal protein S3	1.93e+01
28	63	31.7	1376	1	VCBED6	major capsid protein	1.93e+01
29	62	31.2	345	2	T02295	hypothetical protein	2.72e+01
30	62	31.2	419	2	D70163	hypothetical protein	2.72e+01
31	62	31.2	527	1	S25478	heat shock transcript	2.72e+01
32	62	31.2	527	1	S25481	heat shock transcript	2.72e+01
33	62	31.2	572	1	HNN273	hemagglutinin-neurami	2.72e+01
34	62	31.2	572	1	HNN274	hemagglutinin-neurami	2.72e+01
35	62	31.2	572	1	HNN279	hemagglutinin-neurami	2.72e+01
36	62	31.2	572	1	HNN279	hemagglutinin-neurami	2.72e+01
37	62	31.2	572	1	HNN283	hemagglutinin-neurami	2.72e+01
38	62	31.2	572	2	A46451	hemagglutinin-neurami	2.72e+01
39	62	31.2	572	1	HNN2P3	hemagglutinin-neurami	2.72e+01
40	62	31.2	572	1	HNN282	hemagglutinin-neurami	2.72e+01
41	62	31.2	617	2	S77077	hypothetical protein	2.72e+01
42	62	31.2	714	2	A41464	hemolysin secretion p	2.72e+01
43	62	31.2	901	1	RGEENT	regulatory protein ma	2.72e+01
44	62	31.2	997	2	A40812	Ca2+-transporting ATP	2.72e+01
45	62	31.2	1042	2	B40812	Ca2+-transporting ATP	2.72e+01

ALIGNMENTS

RESULT 1
 ENTRY A45913 #type complete
 TITLE plantaricin A - Lactobacillus plantarum
 ORGANISM #formal_name Lactobacillus plantarum
 DATE 14-May-1993 #sequence_revision 14-May-1993 #text_change 12-Sep-1997
 ACCESSIONS A45913
 REFERENCE A45913
 #authors Nissen-Meyer, J.; Granly-Larsen, A.; Sletten, K.; Daeschel, M.; Nes, I.F.
 #submission submitted to the Protein Sequence Database, April 1993
 #accession A45913
 #status preliminary
 #molecule_type protein
 #residues 1-22 ##label NIS
 #length 22 #molecular-weight 2497 #checksum 9237
 KEYWORDS antibiotic; bacteriocin
 SUMMARY #length 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Query Match 83.4%; Score 166; DB 2; Length 22;
 Best Local Similarity 100.0%; Pred. No. 5.88e-20;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AYSIQMGATAIKOVKKLFKKWG 22
 QY 4 AYSIQMGATAIKOVKKLFKKWG 25
 RESULT 2
 ENTRY S25345 #type complete
 TITLE probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)
 ALTERNATE_NAMES hypothetical protein YCR1102
 ORGANISM #formal_name Saccharomyces cerevisiae
 DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 06-Feb-1998
 ACCESSIONS S25345; S19504
 REFERENCE S25345
 #authors Wilson, C.; Grisanti, P.; Frontali, L.
 #journal Yeast (1992) 8:569-575
 #title The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromosome III contains two new open reading frames.
 #cross-references MUID:92397594
 #accession S25345
 #molecule_type DNA
 #residues 1-1609 #label WIL
 ##cross-references GB:X59720; EMBL:S43845; NID:g1907116; PID:e264634;

```

REFERENCE S19504 PID:g1907227
#authors Frontali, L.; Grisanti, P.
#submission Submitted to the Protein Sequence Database, March 1992
#accession S19504
#molecule_type DNA
#residues 1-1609 #label FRO
#cross-references EMBL:X59720; NID:g19071116; PID:e264634; PID:g1907227;
MIPS:YCR089W

GENETICS
#gene SGD:FIG2
#map_position 3R
#cross-references SGD:S0000685; MIPS:YCR089W
KEYWORDS transmembrane protein
FEATURE
4-20 #domain transmembrane #status predicted #label TM1\
1592-1609 #domain transmembrane #status predicted #label TM2
SUMMARY #length 1609 #molecular-weight 166047 #checksum 8346
Query Match 39.2%; Score 78; DB 2; Length 1609;
Best Local Similarity 34.8%; Pred. No. 7.06e-02;
Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;
Db 1092 PSQYSLSTATTINGIKTVTTW 1114
QY 2 SSAYSQMGATAIKQVKLFKKW 24

RESULT 3
ENTRY A42420 #type complete
TITLE L-iduronidase (EC 3.2.1.76) - dog
ORGANISM #formal_name Canis lupus familiaris #common_name dog
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
29-Jan-1999
ACCESSIONS A42420
REFERENCE A42420
#authors Stoltzfus, L.J.; Sosa-Pineda, B.; Moskowitz, S.M.; Menon, K.P.; Dlott, B.; Hooper, L.; Teplow, D.B.; Shull, R.M.; Neufeld, E.F.
#journal J. Biol. Chem. (1992) 267:6570-6575
#title Cloning and characterization of cDNA encoding canine alpha-L-iduronidase. mRNA deficiency in mucopolysaccharidosis I dog.
#cross-references MUID:92202199
#accession A42420
#status preliminary
#molecule_type mRNA
#residues 1-655 #label STO
#cross-references GB:M81893; NID:g163963; PID:g163964
#experimental_source testis
#note sequence extracted from NCBI backbone (NCBIP:89830)
CLASSIFICATION #superfamily L-iduronidase
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 655 #molecular-weight 72939 #checksum 4797
Query Match 38.2%; Score 76; DB 2; Length 655;
Best Local Similarity 33.3%; Pred. No. 1.56e-01;
Matches 8; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
Db 267 SSIYILEQEQATVQIIRRLFPKFA 290
QY 2 SSAYSQMGATAIKQVKLFKKW 25

RESULT 4
ENTRY D71021 #type complete
TITLE hypothetical protein PH1465 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998
ACCESSIONS D71021
REFERENCE A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;

```

```

Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Negai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession D71021
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-650 #label KAW
#cross-references GB:AP000066; NID:g3236133; PID:d1031515; PID:g3257889
#experimental_source strain OT3
#note this accession replaces an interim accession for a
sequence replaced by GenBank
GENETICS
#gene PH1465
SUMMARY #length 650 #molecular-weight 75553 #checksum 3311
Query Match 37.7%; Score 75; DB 2; Length 650;
Best Local Similarity 58.3%; Pred. No. 2.30e-01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 615 VKIKKIFERKKG 626
QY 14 IKQVKLFKKWG 25

RESULT 5
ENTRY JEO356 #type complete
TITLE gamma-aminobutyric acid receptor B precursor - human
ALTERNATE_NAMES GABA(B) receptor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change
12-Feb-1999
ACCESSIONS JEO356
REFERENCE JEO356
#authors Grifa, A.; Totaro, A.; Rommens, J.M.; Carella, M.; Roetto, A.; Borgato, L.; Zelante, L.; Gasparini, P.
#journal Biochem. Biophys. Res. Commun. (1998) 250:240-245
#title GABA (gamma-amino-butyric acid) neurotransmission:
identification and fine mapping of the human GABAB receptor
gene.
#cross-references MUID:98440782
#accession JEO356
#molecule_type mRNA
#residues 1-960 #label GRI
#cross-references GB:Y11044; NID:g2826760
#note This ORF is not annotated in GenBank entry HSGTHLAL,
release 109
GENETICS
#map_position 6p21.3-6p21.3
KEYWORDS glycoprotein; neurotransmitter receptor; transmembrane
protein
FEATURE
1-11 #domain signal sequence #status predicted #label SIG\
12-960 #product gamma-aminobutyric acid receptor B #status
predicted #label MAT\
590-613 #domain transmembrane #status predicted #label TM1\
627-654 #domain transmembrane #status predicted #label TM2\
666-687 #domain transmembrane #status predicted #label TM3\
709-730 #domain transmembrane #status predicted #label TM4\
767-788 #domain transmembrane #status predicted #label TM5\
803-825 #domain transmembrane #status predicted #label TM6\
831-856 #domain transmembrane #status predicted #label TM7\
23,83,439,481,501,
513,630 #binding_site carbohydrate (Asn) (covalent) #status
predicted

```

US-09-068-507A-1.1pr

We Sep 15 12:54:01 1999

title

Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees.

cross-references MUID:95020537

accession S39693

molecule_type DNA

residues 1-649 #label GLA

cross-references EMBL:X73124; NID:9413923; PID:g413962

authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gagliardi, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Hage, K.; Halech, J.; Harwood, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Hosono, S.; C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Karamata, D.; Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Kobayashi, K.; Koster, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Leber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, E.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Tanaka, T.; Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Uchiyama, S.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

journal Nature (1997) 390:249-256

title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

cross-references MUID:98044033

accession P69687

status preliminary; nucleic acid sequence not shown;

molecule_type DNA

residues 1-649 #label KUN

cross-references GB:299123; GB:AL009126; NID:g2636240; PID:el186315;

experimental_source strain 168

GENETICS qoxB

gene heterotetramer; chains I, II, III and IV

complex #superfamily cytochrome-c oxidase chain I; cytochrome-c

classification oxidase chain I homology

keywords chromoprotein; copper binding; electron transfer; heme;

heterotetramer; membrane-associated complex;

oxidoreductase; transmembrane protein

FEATURE

21-37

52-498

57-73

107-123

142-158

192-208

233-249

283-299

domain transmembrane #status predicted #label TM1\

domain cytochrome-c oxidase chain I homology #label

COI\

domain transmembrane #status predicted #label TM2\

domain transmembrane #status predicted #label TM3\

domain transmembrane #status predicted #label TM4\

domain transmembrane #status predicted #label TM5\

domain transmembrane #status predicted #label TM6\

domain transmembrane #status predicted #label TM7\

SUMMARY

length 960 #molecular-weight 108148 #checksum 3766

Query Match 37.7%; Score 75; DB 2; Length 960;

Best Local Similarity 87.5%; Pred. No. 2.30e-01;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 298 KLFKKGW 305

Qy 19 KLFKKGW 26

RESULT 6

ENTRY A40361 #type complete

TITLE virC-region hypothetical protein yscA - Versinia

ORGANISM enterocolitica plasmid pYV

DATE 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change

09-Sep-1997

ACCESSIONS A40361

REFERENCE Michiels, T.; Vanooteghem, J.C.; Lambert de Rouvoit, C.;

#authors China, B.; Gustin, A.; Boudry, P.; Cornet, G.R.

#journal J. Bacteriol. (1991) 173:4994-5009

#title Analysis of virC, an operon involved in the secretion of Yop

proteins by Versinia enterocolitica.

cross-references MUID:91317716

accession A40361

status preliminary

molecule_type DNA

residues 1-32 #label MIC

cross-references GB:M74011; NID:g155549; PID:g155550

GENETICS plasmid

length 32 #molecular-weight 3915 #checksum 9835

Query Match 35.2%; Score 70; DB 2; Length 32;

Best Local Similarity 31.3%; Pred. No. 1.55e-00;

Matches 5; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 1 MSQITTKHITVLFRRM 16

Qy 9 MGATAIKQVKLFKKW 24

RESULT 7

ENTRY B38129 #type complete

TITLE bo-type ubiquinol oxidase (EC 1.10.3.-) chain I - Bacillus

subtilis

ALTERNATE_NAMES cytochrome aa3 quinol oxidase chain I; quinol oxidase aa3-600

ORGANISM #formal_name Bacillus subtilis

DATE 04-Mar-1993 #sequence_revision 15-Oct-1994 #text_change

17-Mar-1999

ACCESSIONS B38129; S39693; F69687

REFERENCE Santaana, M.; Kunst, F.; Hulio, M.F.; Rapoport, G.; Danchin,

#authors A.; Glaser, P.

#journal J. Biol. Chem. (1992) 267:10225-10231

#title Molecular cloning, sequencing, and physiological

characterization of the qox operon from Bacillus subtilis

encoding the aa3-600 quinol oxidase.

cross-references MUID:92268053

accession B38129

molecule_type DNA

residues 1-649 #label SAN

cross-references GB:M86548; NID:g143395; PID:g143397

note sequence extracted from NCBI backbone (NCBIN:103632,

NCBIP:103603)

REFERENCE S39655

#authors Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales,

W.; Hulio, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino,

L.; Moszer, I.; Presecan, E.; Santaana, M.; Schneider, E.;

Schweizer, J.; Vertes, A.; Rapoport, G.; Danchin, A.

Mol. Microbiol. (1993) 10:371-384

#journal

Accession C71060

```

#length 341 #checksum 822
    try Match          33.78; Score 67; DB 2; Length 341;
    Local Similarity 46.28; Pred. No. 4.68e+00;
    6; Conservative    5; Mismatches 2; Indels 0; Gaps 0;

```

US-09-068-507A-1.rpr

Wed Sep 15 12:54:01 1999

```

Db 133 SOVKSLRKFKW 145
QY 12 TAIKQVKLFK 24

RESULT 12
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors

A71174 #type complete
hypothetical protein PH0589 - Pyrococcus horikoshii
#formal_name Pyrococcus horikoshii
14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998

A71174
A71000
Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Yasuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references NID:98344137
#accession A71174
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-457 #label KAW
#cross-references GB:AP000002; NID:g3236129; PID:d1030621; PID:g3256995
#experimental_source strain OT3
#note this accession replaces an interim accession for a
sequence replaced by GenBank

GENETICS
#gene
SUMMARY
Query Match 33.7%; Score 67; DB 2; Length 457;
Best Local Similarity 41.7%; Pred. No. 4.68e+00; Indels 0; Gaps 0;
Matches 5; Conservative 4; Mismatches 3;
Db 97 REVREYMRKKGW 108
QY 15 KQVKLFKKGW 26

RESULT 13
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors

F69771 #type complete
conserved hypothetical protein ydbk - Bacillus subtilis
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998

F69771
A69580
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devigne, K.M.; Duesterhoeft, A.;
Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmata, D.;
Kashara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Medigue, C.; Medina, N.; Meilado, R.P.; Mizuno,
Maueel, C.; Nakai, S.; Noback, M.; Moore, S.H.; Parro,
M.; Moestl, D.; Ogiwara, A.; Oudega, B.; Park, Prescott, G.;
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Rapoport, G.;
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rev, M.; Reynolds, S.; Rieger, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, K.; Seror, P.; Shin, B.S.; Solido,
Sekowska, A.; Seror, S.J.; Serror, P.; Takagi, T.; Tanaka, T.;
B.; Sorokin, A.; Tacconi, E.; Tamakoshi, A.; Uchiyama, S.;
Takemaru, K.; Takeuchi, M.; Tashiro, V.; Uchiyama, S.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references NID:98044033
#accession F69771
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-246 #label KUN
#cross-references GB:299106;
PID:g2832750
#experimental_source strain 168

GENETICS
#gene
SUMMARY
Query Match 33.2%; Score 66; DB 2; Length 246;
Best Local Similarity 52.9%; Pred. No. 6.70e+00; Indels 1; Gaps 1;
Matches 9; Conservative 6; Mismatches 1;
Db 175 GASTVLQMLKLFKNG 191
QY 10 GATAIKQVKLF-KKWG 25

RESULT 14
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors

I52374 #type complete
Cerebrin-50 - human
#formal_name Homo sapiens #common_name man
02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
02-Jul-1996

I52374
Li, A.H.; Silvestrini, B.; Leone, M.G.; Giacomelli, S.;
Cheng, C.Y.
Biochem. Mol. Biol. Int. (1995) 35:135-144
Cerebrin-50, a human cerebrospinal fluid protein whose mRNA
is present in multiple tissues but predominantly expressed
in the lymphoblastoid cells and the brain.
#cross-references NID:95253026

#cross-references NID:95253026
#accession I52374
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-435 #label RES
#cross-references GB:S76853; NID:g913522; PID:g913523
#length 435 #molecular-weight 51487 #checksum 3663

SUMMARY
Query Match 33.2%; Score 66; DB 2; Length 435;
Best Local Similarity 41.7%; Pred. No. 6.70e+00;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
Db 74 SVHSLQKNGEPKQKQFIKTDW 97
QY 3 SAYSLQMGATAIKQVKLFKKGW 26

```

```

RESULT 15
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:94170813
#accession S42208
#molecule_type mRNA
##residues_type protein
##cross-references EMBL:D15482; NID:9473742; PID:d100459; PID:9538248
#accession S71496
#molecule_type protein
##residues 170-189:721-736:813-819:879-885 #label MAX
CLASSIFICATION #superfamily NAD+ ADP-ribosyltransferase
KEYWORDS DNA binding; glycosyltransferase; NAD; pentosyltransferase;
zinc finger
FEATURE
1-369 #domain DNA binding #status predicted #label DNA
370-507 #domain auto-modification #status predicted #label NAD
508-996 #domain NAD binding #status predicted #label AMO\
SUMMARY #length 996 #molecular-weight 113018 #checksum 1061
Query Match 33.2% Score 66; DB 2; Length 996;
Best Local Similarity 34.6%; Pred.No. 6.70e+00;
Matches 9; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
Db 552 KNSFYKLQLESDMKNRWVRSWGR 577
Oy 1 KSSAYSLOMGATAIKQVKLFKKWGW 26

```

Search completed: Wed Sep 15 10:38:41 1999
Job time : 17 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:38:57 1999; MasPar time 3.11 Seconds
Tabular output not generated. 236.414 Million cell updates/sec

Title: >US-09-068-507A-1
Description: (1-26) from US09068507A.pep
Perfect Score: 199
Sequence: 1 KSSAYSLONGATAIKQVKKLFKKGW 26

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 31.897; Variance 43.884; scale 0.727

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	199	100.0	48	1	PLNA_LACPL	5.90e-32
2	78	39.2	1609	1	FIG2_YEAST	1.20e-02
3	76	38.2	655	1	IDUA_CANFA	2.93e-02
4	70	35.2	32	1	YSCA_YEREN	3.91e-01
5	69	34.7	649	1	QOX1_BACSU	5.93e-01
6	68	34.2	134	1	Y007_MYCLE	8.97e-01
7	67	33.7	280	1	EAEA_HAFAL	1.35e+00
8	67	33.7	341	1	YHS3_SACKL	1.35e+00
9	66	33.2	996	1	PPOL_SARPE	2.02e+00
10	65	32.7	419	1	KCRS_CHICK	3.02e+00
11	64	32.2	529	1	YBBB_BACSU	4.48e+00
12	63	31.7	1376	1	VCAP_HSVEB	9.73e+00
13	62	31.2	102	1	YMI4_YEAST	9.73e+00
14	62	31.2	221	1	GTAL_RAT	9.73e+00
15	62	31.2	527	1	HSF8_LYCPE	9.73e+00
16	62	31.2	572	1	HMA_P13HW	9.73e+00
17	62	31.2	572	1	HMA_P13HA	9.73e+00
18	62	31.2	572	1	HMA_P13HU	9.73e+00
19	62	31.2	572	1	HMA_P13HV	9.73e+00
20	62	31.2	572	1	HMA_P13HT	9.73e+00
21	62	31.2	572	1	HMA_P13HX	9.73e+00
22	62	31.2	572	1	HMA_P13HD	9.73e+00
23	62	31.2	617	1	VC31_BPMD2	9.73e+00

ID	PLNA_LACPL	STANDARD;	PRT;	48 AA.		
24	62	31.2	901	1	MALT_ECOLI	9.73e+00
25	62	31.2	1131	1	YAB9_YEAST	9.73e+00
26	61	30.7	196	1	RETB_CHICK	1.42e+01
27	61	30.7	209	1	UBIG_SALTY	1.42e+01
28	61	30.7	218	1	Y007_MYCTU	1.42e+01
29	61	30.7	231	1	PFS_BACSU	1.42e+01
30	61	30.7	240	1	UBIG_ECOLI	1.42e+01
31	61	30.7	418	1	KCRU_MOUSE	1.42e+01
32	61	30.7	439	1	YGFP_ECOLI	1.42e+01
33	61	30.7	454	1	GLNA_METJA	1.42e+01
34	61	30.7	572	1	HEMA_P13B	1.42e+01
35	61	30.7	724	1	LEM2_CANFA	1.42e+01
36	61	30.7	724	1	VG01_BPP22	1.42e+01
37	61	30.7	766	1	DOC3_MOUSE	1.42e+01
38	61	30.7	770	1	DOC2_HUMAN	1.42e+01
39	61	30.7	1234	1	JMJ_MOUSE	1.42e+01
40	60	30.2	170	1	Y531_METJA	2.07e+01
41	60	30.2	406	1	YC08_METJA	2.07e+01
42	60	30.2	466	1	DC0R_YEAST	2.07e+01
43	60	30.2	585	1	YEJM_HAEIN	2.07e+01
44	60	30.2	785	1	YE15_CAEEL	2.07e+01
45	60	30.2	1064	1	YY08_METJA	2.07e+01

ALIGNMENTS

RESULT 1
ID PLNA_LACPL STANDARD; PRT; 48 AA.
AC P80214;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BACTERIOCIN PLANTARICIN A PRECURSOR.
GN PLNA.
OS LACTOBACILLUS PLANTARUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;
OC LACTOBACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C11;
RX MEDLINE: 94161498.
RA DIEP B.D., HAVARSTEIN L.S., NISSEN-MEYER J., NES F.I.;
RT "The gene encoding plantaricin A, a bacteriocin from Lactobacillus plantarum C11, is located on the same transcription unit as an agr-like regulatory system";
RL APPL. ENVIRON. MICROBIOL. 60:160-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C11;
RX MEDLINE: 96345611.
RA DIEP D.B., HAVARSTEIN L.S., NES I.F.;
RT "Characterization of the locus responsible for the bacteriocin production in Lactobacillus plantarum C11";
RL J. BACTERIOL. 178:4472-4483(1996).
RN [3]
RP SEQUENCE OF 26-47.
RC STRAIN-C11;
RX MEDLINE: 94065628.
RA NISSEN-MEYER J., GRANLY LARSEN A.G., SLETTEN K., DAESCHEL M., NES I.F.;
RT "Purification and characterization of plantaricin A, a Lactobacillus plantarum bacteriocin whose activity depends on the action of two peptides";
RL J. GEN. MICROBIOL. 139:1973-1978(1993).
CC -!- FUNCTION: THIS HEAT STABLE BACTERIOCIN INHIBITS THE GROWTH OF CLOSELY RELATED LACTOBACILLUS SPECIES. IT MAY ACT AS A PORE-FORMING PROTEIN, CREATING A CHANNEL IN THE CELL MEMBRANE THROUGH A "BARREL STAVE" MECHANISM.
CC -!- SUBUNIT: ACTIVE PLANTARICIN A IS COMPOSED OF AN ALPHA CHAIN AND A BETA CHAIN
CC -!- THE BETA CHAIN SEQUENCE IS SHOWN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X75323; G452406; -;
 DR EMBL; X94434; E217591; -;
 DR PIR; A45913; A45913.
 KW ANTI-BIOTIC; BACTERIOICIN.
 FT PROPEP 1 25
 FT CHAIN 26 48 BACTERIOICIN PLANTARICIN A.
 FT VARIANT 26 26 MISSING (IN THE ALPHA CHAIN).
 FT SEQUENCE 48 AA; 5458 MW; A6083DE9 CRC32;

Query Match 100.0%; Score 199; DB 1; Length 48;
 Best Local Similarity 100.0%; Pred. No. 5.90e-32;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 23 KSSAYSLQMGATAIKQVKLFKKGW 48
 QY 1 KSSAYSLQMGATAIKQVKLFKKGW 26

RESULT 2
 ID FIG2 YEAST STANDARD; PRT: 1609 AA.

AC P25653;
 DT 01-MAY-1992 (REL. 22, CREATED)
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE FACTOR INDUCED GENE 2.

GN FIG2 OR YCR089W OR YCR89W OR YCR1102.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.

[1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 92397594.

RA WILSON C., GRISANTI P., FRONTALI L.;

RT "The complete sequence of a 6146 bp fragment of Saccharomyces

cerevisiae chromosome III contains two new open reading frames.";
 RL YEAST 8:569-575(1992).

CC -1- FUNCTION: REQUIRED FOR EFFICIENT MATING.

CC -1- INDUCTION: BY MATING PHEROMONES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X59720; E264634; -;
 DR PIR; S19504; S19504.
 DR PIR; S25345; S25345.
 DR SGD; L000312; FIG2.
 SQ SEQUENCE 1609 AA; 166049 MW; DE974CE8 CRC32;

Query Match 39.2%; Score 78; DB 1; Length 1609;
 Best Local Similarity 34.8%; Pred. No. 1.20e-02;
 Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 1092 PSQYSLSTATTINGIKTVYTTW 1114

QY 2 SSAYSLQMGATAIKQVKLFKKW 24

RESULT 3

ID IDUA CANFA STANDARD; PRT: 655 AA.

AC Q01634;

DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE ALPHA-L-IDURONIDASE PRECURSOR (EC 3.2.1.76).
 GN IDUA.

OS CANIS FAMILIARIS (DOG).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.

[1]
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RP TISSUE-TESTIS;

RC MEDLINE; 92202199.

RA STOLTZFUS L.J., SOSA-PINEDA B., MOSKOWITZ S.M., MENON K.P., DLOTT B.,
 RA HOOPER L., TEPLow D.B., SHULL R.M., NEUFELD E.F.;

RT "Cloning and characterization of cDNA encoding canine alpha-L-

iduronidase. mRNA deficiency in mucopolysaccharidosis I dog.";

RL J. BIOL. CHEM. 267:6570-6575(1992).

[2]

RN SEQUENCE FROM N.A.

RP TISSUE-TESTIS, AND FIBROBLAST;

RC MEDLINE; 93052413.

RA MENON K.P., TIEU P.T., NEUFELD E.F.;

RT "Architecture of the canine IDUA gene and mutation underlying canine

mucopolysaccharidosis I";

RL GENOMICS 14:763-768(1992).

CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF ALPHA-L-IDURONOSIDIC LINKAGES IN

CC DESULFATED DERMATAN.

CC -1- SUBUNIT: MONOMER (PROBABLE).

CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.

CC -1- TISSUE SPECIFICITY: FOUND UBQUITOUSLY.

CC -1- PTM: A SMALLER 63 KDA PROTEIN PROBABLY ARISES FROM IDUA PROTEIN

CC BY PROTEOLYTIC CLEAVAGE.

CC -1- DISEASE: DEFECTS IN IDUA ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS

CC TYPE I (MPS I).

CC -1- SIMILARITY: BELONGS TO FAMILY 39 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L01058; -; NOT_ANNOTATED_CDS.

DR EMBL; L01059; -; NOT_ANNOTATED_CDS.

DR EMBL; L01060; -; NOT_ANNOTATED_CDS.

DR EMBL; L01061; -; NOT_ANNOTATED_CDS.

DR EMBL; L01065; G52348; -;

DR EMBL; M81893; G163964; -;

PIR; A42420; A42420.

DR PROSITE; PS01027; GLYCOSYL_HYDROL_F39; 1.

DR PFAM; PF01229; GLYCOSYL_hydrl9; 1.

KW HYDROLASE; GLYCOSIDASE; LYSOSOME; SIGNAL.

FT SIGNAL 1 25

FT CHAIN 26 655 ALPHA-L-IDURONIDASE.

FT ACT_SITE 181 181 PROTON DONOR (POTENTIAL).

FT ACT_SITE 298 298 NUCLEOPHILE (POTENTIAL).

FT CARBOHYD 109 109 POTENTIAL.

FT CARBOHYD 189 189 POTENTIAL.

FT CARBOHYD 242 242 POTENTIAL.

FT CARBOHYD 335 335 POTENTIAL.

FT CARBOHYD 371 371 POTENTIAL.

FT CARBOHYD 414 414 POTENTIAL.

SQ SEQUENCE 655 AA; 72939 MW; 7358E9D7 CRC32;

Query Match 38.2%; Score 76; DB 1; Length 655;
 Best Local Similarity 33.3%; Pred. No. 2.93e-02;

Matches 8; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

Db 267 SSIIYLEQEQATVQIRKLPKFA 290

QY 2 SSAYSLQMGATAIKQVKLFKKW 25


```

FT METAL 417 417 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 649 AA; 73838 MW; 7C64B76E CRC32;

Query Match 34.7%; Score 69; DB 1; Length 649;
Best Local Similarity 42.9%; Pred. No. 5.93e-01;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 22 STALSTIAIIFVLTYFKKWKW 42
   [: : : : : | | | | |
QY 6 SLQMGATAIKQVKLFFKKWG 26

RESULT 6
ID YQ07_MYCLE STANDARD; PRT; 134 AA.
AC Q49642;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 14.2 KD PROTEIN BL177_C2_172.
GN BL177_C2_172.
OS MYCOBACTERIUM LEPRAE.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RL ROBLSON K., SMITH D.R.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE DEDA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U00011; G466816;
CC SEQUENCE 134 AA; 14249 MW; 7F516A5 CRC32;

Query Match 34.2%; Score 68; DB 1; Length 134;
Best Local Similarity 28.6%; Pred. No. 8.97e-01;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 91 FPKHFGPGHVALVERLFNRWG 111
   : : : : : | : : : : |
QY 5 YSLQMGATAIKQVKLFFKKWG 25

RESULT 7
ID EAEA_HAFAL STANDARD; PRT; 280 AA.
AC P52869;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE INTIMIN (OUTER MEMBRANE PROTEIN) (ATTACHING AND EFFACING PROTEIN)
DE (FRAGMENT)
GN EAEA OR EAE.
OS HAFNIA ALVEL.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC HAFNIA.
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE; 94222551.
RA FRANKEL G., CANDY D.C.A., EVEREST P., DOUGAN G.;
RT "Characterization of the c-terminal domains of intimin-like proteins
RT of enteropathogenic and enterohemorrhagic Escherichia coli,
RT Citrobacter freundii, and Hafnia alvei."
RL INFECT. IMMUN. 62:1835-1842(1994).
CC -!- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS.
CC -!- SUBCELLULAR LOCATION: OUTER SURFACE.

```

```

CC -!- SIMILARITY: BELONGS TO THE EAE/INVASIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L29509; G472359;
CC KW OUTER MEMBRANE.
CC FT NON_TER 1
CC SQ SEQUENCE 280 AA; 30146 MW; FBC2C19D CRC32;

Query Match 33.7%; Score 67; DB 1; Length 280;
Best Local Similarity 40.0%; Pred. No. 1.35e+00;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 209 PSSIKELKLDYDDWG 223
   : : : : : | : :
QY 11 ATAIKQVKLFFKKWG 25

RESULT 8
ID YHS3_SACKL STANDARD; PRT; 341 AA.
AC Q03000;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IN HIS3 3' REGION (FRAGMENT).
OS SACCHAROMYCES KLUYVERI (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CBS 3082;
RX MEDLINE; 93289813.
RA WEINSTOCK K.G., STRATHERN J.N.;
RT "Molecular genetics in Saccharomyces kluyveri: the HIS3 homolog and
RT its use as a selectable marker gene in S. kluyveri and Saccharomyces
RT cerevisiae."
RL YEAST 9:351-361(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z14125; E1175791;
CC DR PIR; S31236; S31236.
CC KW HYPOTHETICAL PROTEIN.
CC FT NON_TER 1
CC FT DOMAIN 103 107 POLY-SER.
CC SQ SEQUENCE 341 AA; 37381 MW; 64772A58 CRC32;

Query Match 33.7%; Score 67; DB 1; Length 341;
Best Local Similarity 46.2%; Pred. No. 1.35e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 133 SQVSKLRKRVKKW 145
   : : : : : | : :
QY 12 TAIKQVKLFFKKW 24

RESULT 9
ID PPOL_SARPE STANDARD; PRT; 996 AA.
AC Q11208;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

```

01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE POLY (ADP-RIBOSE) POLYMERASE (EC 2.4.2.30) (PARP) (ADPRT) (NAD(+)) ADP-
 DE RIBOSYLTRANSFERASE) (POLY(ADP-RIBOSE) SYNTHETASE).
 OS SARCOPHAGA PREGRIINA (FLESH FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; OESTROIDEA;
 OC SARCOPHAGIDAE; SARCOPHAGA.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 94170813.
 RA MASUTANI M., NOZAKI T., HITOMI Y., IKEJIMA M., NAGASAKI K.,
 RA DE PRATI A.C., KURATA S., NATORI S., SUGIMURA T., ESUMI H.;
 RT "Cloning and functional expression of poly(ADP-ribose) polymerase
 RT cDNA from *Sarcophaga peregrina*.";
 RL EUR. J. BIOCHEM. 220:607-614(1994).
 CC [1]
 CC [1]- FUNCTION: POLY(ADP-RIBOSE) POLYMERASE MODIFIES VARIOUS NUCLEAR
 CC PROTEINS BY POLY(ADP-RIBOSYL)ATION. THE MODIFICATION IS DEPENDENT
 CC ON DNA AND IS INVOLVED IN THE REGULATION OF VARIOUS IMPORTANT
 CC CELLULAR PROCESSES SUCH AS DIFFERENTIATION, PROLIFERATION, AND
 CC TUMOR TRANSFORMATION AND ALSO IN THE REGULATION OF THE MOLECULAR
 CC EVENTS INVOLVED IN THE RECOVERY OF CELL FROM DNA DAMAGE.
 CC [1]- CATALYTIC ACTIVITY: NAD(+) + (ADP-D-RIBOSYL)[N]-ACCEPTOR -
 CC NICOTINAMIDE + (ADP-D-RIBOSYL)[N-1]-ACCEPTOR.
 CC [1]- COFACTOR: ZINC, CONTAINS TWO MOLE OF ZINC PER MOLE OF PROTEIN.
 CC [1]- SUBCELLULAR LOCATION: NUCLEAR.
 CC [1]- THE ADP-D-RIBOSYL GROUP OF NAD(+) IS TRANSFERRED TO AN ACCEPTOR
 CC CARBOXYL GROUP ON A HISTONE OR THE ENZYME ITSELF, AND FURTHER ADP-
 CC RIBOSYL GROUPS ARE TRANSFERRED TO THE 2'-POSITION OF THE TERMINAL
 CC ADENOSINE MOIETY, BUILDING UP A POLYMER WITH AN AVERAGE CHAIN
 CC LENGTH OF 20-30 UNITS.
 CC [1]- SIMILARITY: BELONGS TO THE PARP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D16482; G538248; -
 DR PROSITE: PS00347; PARP_ZN_FINGER_1; FALSE_NEG.
 DR PROSITE: PS50064; PARP_ZN_FINGER_2; 2.
 DR PFAM: PF00533; BRCT; 1.
 DR PFAM: PF00644; PARP; 1.
 DR PFAM: PF00645; zf-PARP; 2.
 DR HSP: P26446; 4PAX.
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; NAD; DNA-BINDING; NUCLEAR PROTEIN;
 KW ADP-RIBOSYLATION; ZINC-FINGER; ZINC.
 FT DNA_BIND 1 369 BY SIMILARITY.
 FT DOMAIN 370 507 AUTOMODIFICATION DOMAIN.
 FT DOMAIN 508 996 NAD-BINDING.
 FT ZN_FING 19 54 NAD-BINDING.
 FT ZN_FING 126 164 BY SIMILARITY.
 FT DOMAIN 211 214 NUCLEAR LOCALIZATION SIGNAL 1ST PART.
 FT DOMAIN 232 235 NUCLEAR LOCALIZATION SIGNAL 2ND PART.
 FT ACT_SITE 878 878 BY SIMILARITY.
 FT ACT_SITE 996 AA; 113018 MW; C9E71E28 CRC32;
 SQ SEQUENCE 996 AA; 113018 MW; C9E71E28 CRC32;
 Query Match 33.2%; Score 66; DB 1; Length 996;
 Best Local Similarity 34.6%; Pred. No. 2.02e+00;
 Matches 9; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Db 552 KNSFYKLQLESMDKRNFWVFRSGR 577
 QY 1 KSSAYSLQMGATAIKOVKKLFKKWG 26
 RESULT 10
 ID KCRS_CHICK STANDARD; PRT; 419 AA.
 AC F11009;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE CREATINE KINASE, SARCOMERIC MITOCHONDRIAL PRECURSOR (EC 2.7.3.2) (S-
 DE MTCK) (MIB-CK) (BASIC-TYPE MITOCHONDRIAL CREATINE KINASE).
 GN CKMT2.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE OF 1-39 FROM N.A.
 RC STRAIN-WHITE LEGHORN;
 RX MEDLINE: 96216114.
 RA WUEHLEBACH S.M., WIRZ T., BRAENDLE U., PERRIARD J.-C.;
 RT "Evolution of the creatine kinases. The chicken acidic type
 RT mitochondrial creatine kinase gene as the first nonmammalian gene.";
 RL J. BIOL. CHEM. 271:11920-11929(1996).
 CC [2]
 CC SEQUENCE OF 21-419 FROM N.A., AND SEQUENCE OF 40-69.
 CC TISSUE-SKELETAL MUSCLE, AND HEART;
 RX MEDLINE: 88162838.
 RA HOSSELE J.P., SCHLEGEL J., WEGMANN G., WYSS M., BOEHLEN P.,
 RA EPPENBERGER H.M., WALLIMANN T., PERRIARD J.-C.;
 RT "Distinct tissue specific mitochondrial creatine kinases from chicken
 RT brain and striated muscle with a conserved CK framework.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 151:408-416(1988).
 RN [3]
 RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RP MEDLINE: 96256676.
 RX PRITZ-WOLF K., SCHNYDER T., WALLIMANN T., KABSCH W.;
 RA "Structure of mitochondrial creatine kinase.";
 RL NATURE 381:341-345(1996).
 CC [1]- FUNCTION: REVERSIBLY CATALYZES THE TRANSFER OF PHOSPHATE BETWEEN
 CC ATP AND VARIOUS PHOSPHOGENS (E.G. CREATINE PHOSPHATE). CREATINE
 CC KINASE ISOENZYMES PLAY A CENTRAL ROLE IN ENERGY TRANSDUCTION IN
 CC TISSUES WITH LARGE, FLUCTUATING ENERGY DEMANDS, SUCH AS SKELETAL
 CC MUSCLE, HEART, BRAIN, AND SPERMATOZOEA.
 CC [1]- CATALYTIC ACTIVITY: ATP + CREATINE = ADP + PHOSPHOCREATINE.
 CC [1]- SUBUNIT: EXISTS AS AN OCTAMER COMPOSED OF FOUR MTCK HOMODIMERS.
 CC [1]- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE; OUTER SIDE.
 CC [1]- MITOCHONDRIAL CREATINE KINASE BINDS CARDIOLIPIN.
 CC [1]- TISSUE SPECIFICITY: EXPRESSED IN THE LEG MUSCLE AND HEART.
 CC [1]- SIMILARITY: THE FOUR DIFFERENT ISOZYMES OF CREATINE KINASES: B, M,
 CC AND TWO MITOCHONDRIAL FORMS: ARE HIGHLY SIMILAR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M18866; G211526; ALT_FRAME.
 DR PIR: A27708; A27708.
 DR PIR: ICRK; 07-JUL-97.
 DR PROSITE: PS00112; GUANIDO_KINASE; 1.
 DR PFAM: PF00217; ATP_gua_Ptrains; 1.
 KW TRANSFERASE; KINASE; MULTIGENE FAMILY; MITOCHONDRION; TRANSIT PEPTIDE;
 KW 3D-STRUCTURE.
 FT TRANSIT 1 39 MITOCHONDRION.
 FT CHAIN 40 419 CREATINE KINASE, SARCOMERIC
 FT MITOCHONDRIAL.
 FT ACT_SITE 317 317 BY SIMILARITY.
 SQ SEQUENCE 419 AA; 47084 MW; DEED027E CRC32;
 Query Match 32.7%; Score 65; DB 1; Length 419;
 Best Local Similarity 42.9%; Pred. No. 3.02e+00;
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 290 GLKEVERLIKRGW 303
 QY 13 AIKQVKLFKKWG 26

DE (CLASS-ALPHA) (CLONES PGST94 & PGTR261).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84185691.
RA LAI H.-C.J., LI N.-Q., WEISS M.J., REDDY C.C., TU C.-P.D.;
RT "The nucleotide sequence of a rat liver glutathione S-transferase
FT subunit cDNA clone.";
RL J. BIOL. CHEM. 259:5536-5542(1984).
RN [2]
RP SEQUENCE OF 45-196 FROM N.A.
RX MEDLINE: 82075944.
RA KALINAK J.E., TAYLOR J.M.;
RT "Rat glutathione S-transferase. Cloning of double-stranded cDNA and
FT induction of its mRNA.";
RL J. BIOL. CHEM. 257:523-530(1982).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-G.
CC -!- SUBUNIT: HOMODIMER.
CC -!- RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST OF BINARY
CC COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS DESIGNATED:
CC YA (25KD), YB (27KD) AND YC (28KD).
CC -!- YA AND YC SUBUNITS ARE ENCODED BY SEPARATE GENE FAMILIES OR
CC TRANSCRIPTIONAL UNITS. THERE IS EXTENSIVE HOMOLOGIES BETWEEN
CC THOSE TWO FAMILY OF SEQUENCES.
CC -!- THE CLONES PGTR112 AND PGTR38 (SEE SEPARATE ENTRY) AND CLONES
CC PGST94 AND PGTR261 WERE ISOLATED FROM DIFFERENT LITTERS OF
CC SPRAGUE-DAWLEY RATS SUGGESTING THAT THESE ARE EITHER DIFFERENT
CC ALLELES FOR THE SAME CATALYTIC FUNCTIONS OR DIFFERENT MEMBERS OF
CC A CLOSELY RELATED MULTIGENE FAMILY.
CC -!- IN ADDITION TO ITS ENZYMIC ACTIVITY, THE HOMODIMER OF YA CHAINS,
CC CALLED LIGANDIN, BINDS VARIOUS ORGANIC ANIONS, XENOBIOTICS, AND
CC AZOCARCINOGEN DYES. IT IS A CYTOSOLIC PROTEIN FOUND IN MANY
CC MAMMALIAN TISSUES.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: K01931; G204495; -.
CC PIR: A00591; XURTC.
CC PFAM: PF00043; Gluts; 1.
CC HSP: P08263; LGSF.
CC TRANSFERASE; MULTIGENE FAMILY.
FT INIT-MET 0 0
FT CONFLICT 151 151 K -> R (IN PIR DATA BANK).
SQ SEQUENCE 221 AA; 25480 MW; C115D4C3 CRC32;

Query Match 31.2%; Score 62; DB 1; Length 221;
Best Local Similarity 43.8%; Pred. No. 9.73e+00;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 205 LPMDAKQIEEARKIFK 220
QY 7 LMGATAIKQVKLFK 22
| | | | | : | | | |

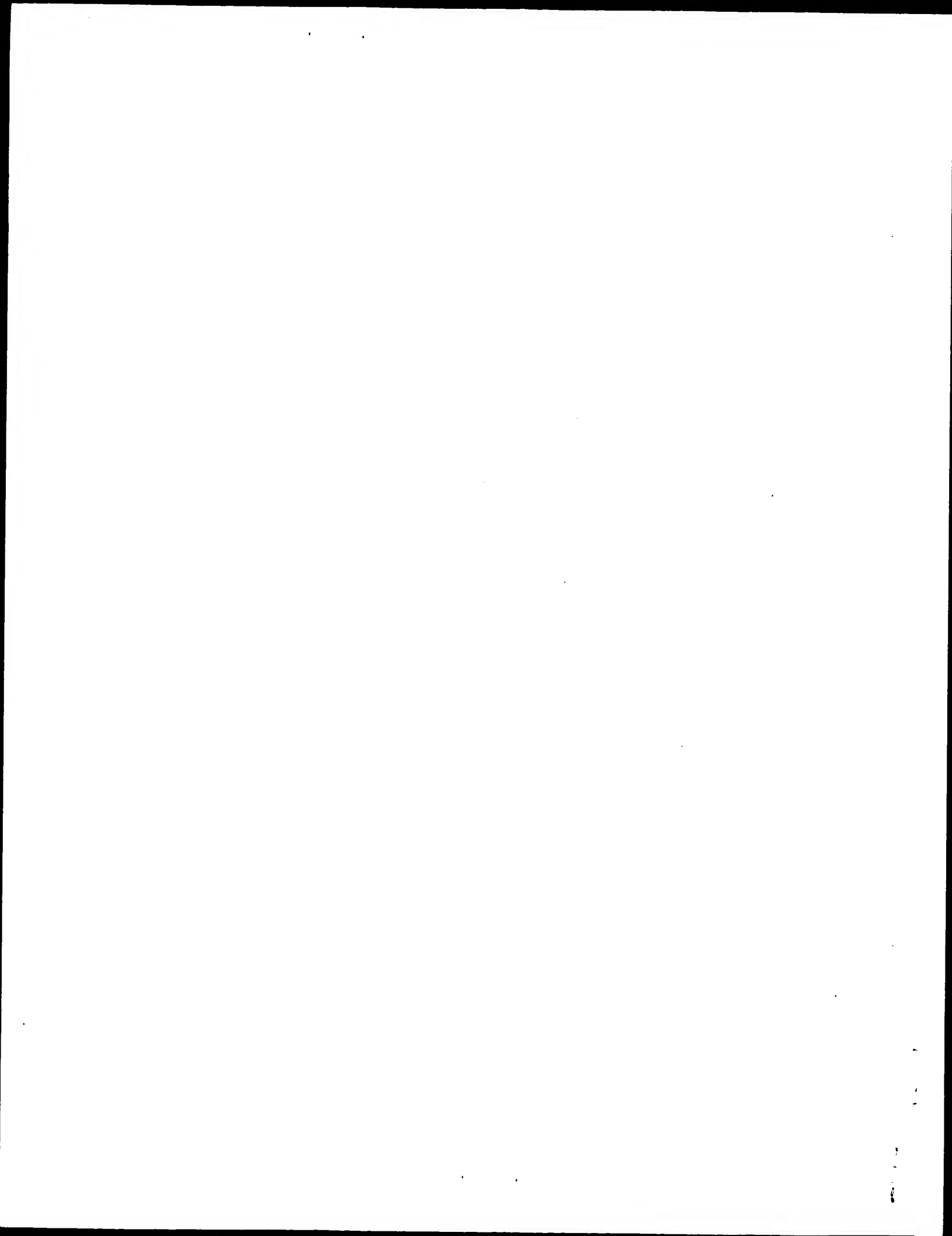
RESULT 15
ID HSF8_LYCPE STANDARD; PRT; 527 AA.
AC P41153;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HEAT SHOCK FACTOR PROTEIN HSF8 (HEAT SHOCK TRANSCRIPTION FACTOR 8)
DE (HSF8) (HEAT STRESS TRANSCRIPTION FACTOR).

GN HSF8.
OS LYCOPERSICON PERUVIANUM (PERUVIAN TOMATO).
OC EUKARYOTA; VIRIDIPLANTAE; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94105354.
RA SCHARF K.D., ROSE S., THIERFELDER J., NOVER L.;
RT "Two cDNAs for tomato heat stress transcription factors.";
RL PLANT PHYSIOL. 102:1355-1356(1993).
CC -!- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
CC PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION (BY
CC SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- PTM: EXHIBITS TEMPERATURE-DEPENDENT PHOSPHORYLATION
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X67600; G19492; -.
CC PIR: S25481; S25481.
CC PROSITE: PS00434; HSF_DOMAIN; 1.
CC PFAM: PF00447; HSF_DNA-Bind; 1.
CC HSP: P21121; 3HSF.
CC TRANSFAC: T02033; -.
CC TRANSCRIPTION REGULATION; NUCLEAR PROTEIN; DNA-BINDING; ACTIVATOR;
KW PHOSPHORYLATION; HEAT SHOCK; MULTIGENE FAMILY.
FT DNA_BIND 39 133 BY SIMILARITY.
FT DOMAIN 14 20 POLY-GLY.
FT DOMAIN 139 143 POLY-GLN.
FT DOMAIN 195 198 POLY-GLN.
SQ SEQUENCE 527 AA; 57519 MW; 34DE58C9 CRC32;

Query Match 31.2%; Score 62; DB 1; Length 527;
Best Local Similarity 64.3%; Pred. No. 9.73e+00;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 514 QMGATNKKQKHKMI 527
QY 8 QMGATAIKQVKLF 21
| | | | | : | | | |

Search completed: Wed Sep 15 10:39:05 1999
Job time : 8 secs.



(TM)

The image shows a row of six dashed-line patterns, each enclosed within a rectangular frame. From left to right, the patterns are:

- A horizontal bar with two vertical bars extending downwards from its center.
- A square-like shape with a small rectangle attached to its top-right side.
- A square-like shape with a diagonal line running from the bottom-left corner to the top-right corner.
- A square-like shape with a small rectangle attached to its top-right side, similar to the second pattern but with slightly different proportions.
- A square-like shape with a small rectangle attached to its top-right side, similar to the previous ones.
- A complex shape consisting of a horizontal bar at the top, a vertical bar on the right, and a zig-zag line forming a triangular shape on the left.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:39:22 1999; Maspar time 5.79 Seconds
245.104 Million cell updates/sec
Irregular output not generated.

```
>US-09-068-507A-1
Title:
Description: (1-26) from US09068507A.pcp
Perfect Score: 199
sequence: 1 KSSAYSLSQMGATAIKQVKLFFKKWG 26
```

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus
```

Mean 30.932; Variance 45.599; scale 0.678

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, based on the distribution determined by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	77	38.7	822	11	054923	RSEC15.	5.45e-02
2	75	37.7	650	11	059134	650AA LONG HYPOTHETICA	1.27e-01
3	75	37.7	844	11	008621	GABA-BR1B RECEPTOR.	1.27e-01
4	75	37.7	960	11	008620	GABA-BR1A RECEPTOR.	2.92e-01
5	73	36.7	950	14	066578	GLYCOPROTEIN GP14.	2.92e-01
6	72	36.2	218	5	Q17890	COSMID C06A8.	4.40e-01
7	69	34.7	450	13	073786	C-SRC KINASE.	1.48e+00
8	69	34.7	1170	10	048573	PUTATIVE DISEASE RESIS	1.48e+00
9	68	34.2	157	1	058919	157AA LONG HYPOTHETICA	2.20e+00
10	68	34.2	256	9	049635	HYPOTHETICAL 23.1 KD P	2.20e+00
11	68	34.2	256	9	038150	PHAGE SP1 DNA SEQUENC	2.20e+00
12	67	33.7	51	13	Q91567	PROTEIN-TYROSINE KINAS	3.26e+00
13	67	33.7	225	14	Q66072	PROTEIN-PROTEIN (FRAGME	3.26e+00
14	67	33.7	457	1	058338	457AA LONG HYPOTHETICA	3.26e+00
15	67	33.7	646	14	Q37319	NUCLEOSIDE TRIPHOSPHAT	3.26e+00
16	67	33.7	912	5	Q19712	F22B5.8.	3.26e+00
17	67	33.7	1415	5	Q61442	ZYG-9.	3.26e+00
18	66	33.2	246	2	P96606	PROBABLE MEMBRANE SPAN	4.80e+00
19	66	33.2	375	4	Q35150	MYELIN TRANSCRIPTION F	4.80e+00
20	66	33.2	435	4	Q16301	CEREBRIN-50.	4.80e+00

21	66	33.2	767	5	Q20176	COSMID F389.	4.80e+00
22	65	32.7	397	5	Q16371	F29G.6 PROTEIN.	7.05e+00
23	65	32.7	492	14	O80624	GAG PROTEIN.	7.03e+00
24	64	32.2	278	2	O86433	UDP-GLUCOSE PYROPHOSPH	1.03e+01
25	64	32.2	370	5	O67139	F22E5.1 PROTEIN.	1.03e+01
26	64	32.2	470	10	O55028	POLLEN-SPECIFIC 1-AMIN	1.03e+01
27	64	32.2	1130	5	O45624	H19N07.2B PROTEIN.	1.03e+01
28	64	32.2	1133	5	O45623	H19N07.2A PROTEIN.	1.03e+01
29	63	31.7	176	5	Q26633	E-ABL MRNA (FRAGMENT).	1.50e+01
30	63	31.7	230	14	O89660	DNA POLYMERASE (FRAGME	1.50e+01
31	63	31.7	265	2	O50975	CONSERVED HYPOTHETICAL	1.50e+01
32	63	31.7	276	11	Q63777	HYPOTHETICAL 32.0 KD P	1.50e+01
33	63	31.7	340	11	Q64175	REVERSE TRANSCRIPTASE	1.50e+01
34	63	31.7	367	1	O29133	ENOLASE (ENO).	1.50e+01
35	63	31.7	367	10	O40993	NTH1.	1.50e+01
36	63	31.7	397	2	O05670	PUTATIVE SERINE PROTEA	1.50e+01
37	63	31.7	419	5	Q24548	T48 MRNA PRECURSOR.	1.50e+01
38	63	31.7	513	11	Q63289	L1 RETROPOSOM, ORF2 MR	1.50e+01
39	63	31.7	556	11	Q63305	LONG INTERSPERSED REPE	1.50e+01
40	63	31.7	599	2	O44107	MAJOR SURFACE PROTEIN	1.50e+01
41	63	31.7	600	3	O74977	HYPOTHETICAL 68.8 KD P	1.50e+01
42	63	31.7	645	11	Q63779	HYPOTHETICAL 76.2 KD P	1.50e+01
43	63	31.7	686	11	Q63288	L1 RETROPOSOM, ORF2 MR	1.50e+01
44	63	31.7	1300	11	P97692	L1 RETROPOSOM ORF	1.50e+01
45	62	31.2	574	14	O89237	HEMAGGLUTININ-NEURAMIN	2.18e+01

ALIGNMENTS

RESULT	1	PRT:	822 AA.
ID	054923	PRELIMINARY;	
AC	054923		
01	JUN-1998	(TREMBLREL. 06, CREATED)	
01	JUN-1998	(TREMBLREL. 06, LAST SEQUENCE UPDATE)	
01	JUN-1998	(TREMBLREL. 06, LAST SEQUENCE UPDATE)	
01	NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE	RSEC15.		
DE	RATUUS NORVEGICUS (RAT).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.		
OC	[1]		
OC	SEQUENCE FROM N.A.		
OC	TISSUE-BRAIN:		
OC	MEDLINE; 98070770.		
OC	KEE Y., YOO J.-S., HAZUKA C.D., PETERSON K.E., HSU S.C.,		
OC	SCHELLER R.H.;		
OC	"Subunit structure of the mammalian exocyst complex. "		
OC	PROC. NATL. ACAD. SCI. U.S.A. 94:14438-14443(1997).		
OC	DEMBL; AF032668; G2827162; "		
OC	SEQUENCE 822 AA; 95105 MW; 38EFF544 CRC32;		

```

Query Match      38.7%  Score 77  DB 11:  Length 822;
Best Local Similarity 33.3%  Pred. No. 5.45e-02;
Matches          7;  Conservative 7;  Mismatches 7;  Indels 0;  Gaps 0;

```

db 731 TLQLAFIDLRQLDLFLMVWDW 751
:|:: ::|: | |
ov 6 SLOMGATAIKOVKKLEKKWG 26

RESULT 2
PRELIMINARY: 650 AA.
PRT;

AC O59134; 01-AUG-1998 (TREMBLER. 07, CREATED)
DT 01-AUG-1998 (TREMBLER. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLER. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLER. 09, LAST ANNOTATION UPDATE)
DT 01-JAN-1999 (TREMBLER. 09, LAST ANNOTATION UPDATE)
EF 550A LONG HYPOTHETICAL PROTEIN.

OS PYROCOCCLUS HORIKOSHII.
ON PH463.
GN
TS
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCACEAE; PYROCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE: 98344137.

RA KAWABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOKAWA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KOSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 3:55-76(1998).
DR EMBL; AP000006; D1031515;
SQ SEQUENCE 650 AA; 75553 MW; EDC92E1E CRC32;
Query Match 37.7%; Score 75; DB 1; Length 650;
Best Local Similarity 58.3%; Pred. No. 1.27e-01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 615 VKKKKIFKKGW 626
QY 14 IKQVKKLFKKGW 25
RESULT 3
ID O08621 PRELIMINARY; PRT; 844 AA.
AC O08621
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE GABA-BRIB RECEPTOR.
DE GABA-BRIB.
GN GABA-BRIB.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIF (RAI F(SPF)); TISSUE=CORTEX, AND CEREBELLUM;
RX MEDLINE; 97222131.
RA KAUPMANN K., HEGGEL K., HEID J., FLOR P.J., BISCHOFF S., MICKEL S.J.,
RA MCMASTER G., ANGST C., BITTIGER H., FROESTL W., BETTLER B.;
RT "Expression cloning of GABA(B) receptors uncovers similarity to
RT metabotropic glutamate receptors.";
RL NATURE 386:239-246(1997).
DR EMBL; Y10370; E311196;
DR PFAM; PF01094; ANF_receptor; 1.
SQ SEQUENCE 844 AA; 95037 MW; 6611F68D CRC32;
Query Match 37.7%; Score 75; DB 11; Length 844;
Best Local Similarity 87.5%; Pred. No. 1.27e-01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 182 KLFEKKGW 189
QY 19 KLFEKKGW 26
RESULT 4
ID O08620 PRELIMINARY; PRT; 960 AA.
AC O08620
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE GABA-BRIB RECEPTOR.
DE GABA-BRIB.
GN GABA-BRIB.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIF(RAI F(SPF)); TISSUE=CORTEX, AND CEREBELLUM;
RX MEDLINE; 97222131.
RA KAUPMANN K., HEGGEL K., HEID J., FLOR P.J., BISCHOFF S., MICKEL S.J.,
RA MCMASTER G., ANGST C., BITTIGER H., FROESTL W., BETTLER B.;
RT "Expression cloning of GABA(B) receptors uncovers similarity to
RT metabotropic glutamate receptors.";
RL NATURE 386:239-246(1997).
DR EMBL; Y10370; E311196;
DR PFAM; PF01094; ANF_receptor; 1.
SQ SEQUENCE 844 AA; 95037 MW; 6611F68D CRC32;

RL NATURE 386:239-246(1997).
DR EMBL; Y10369; E311192;
DR PFAM; PF00084; sushi; 2.
DR PFAM; PF01094; ANF_receptor; 1.
SQ SEQUENCE 960 AA; 108203 MW; CD3A2D9E CRC32;
Query Match 37.7%; Score 75; DB 11; Length 960;
Best Local Similarity 87.5%; Pred. No. 1.27e-01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 298 KLFEKKGW 305
QY 19 KLFEKKGW 26
RESULT 5
ID O66678 PRELIMINARY; PRT; 950 AA.
AC O66678
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE GLYCOPROTEIN GP14.
GN GP14.
OS EQUINE HERPESVIRUS TYPE 1 (EHV-1).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RACH;
RA NEUBAUER A., BRAUN B., KAADEN O.R., OSTERRIEDER K.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X95377; E220895;
DR PFAM; PF00606; Glycoprotein_B; 1.
SQ SEQUENCE 950 AA; 106568 MW; B2DE30B2 CRC32;
Query Match 36.7%; Score 73; DB 14; Length 950;
Best Local Similarity 47.6%; Pred. No. 2.92e-01;
Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 936 MSWVALEKQKRAKKWGW 946
QY 7 LMGATAIKQVKKLFKK-WGW 26
RESULT 6
ID Q17690 PRELIMINARY; PRT; 218 AA.
AC Q17690
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE COSMID C06A8.
GN C06A8.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.

US-09-068-507A-1.rspt

Wed Sep 15 12:54:02 1999

```

RA LEIMBACH D.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U39849; G1055042;
SQ SEQUENCE 218 AA; 25556 MW; 1ACF80C3 CRC32;

Query Match
Best Local Similarity 36.2%; Score 72; DB 5; Length 218;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 106 IATSVKQVEKLFQK 120
QY 9 MGATAIKQVKLFKK 23

RESULT 7 PRELIMINARY; PRT; 450 AA.
ID O73786;
AC O73786;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE C-SRC KINASE.
GN CSK.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA MURPHY S.M., MORGAN D.O.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF052430; G2967840;
SQ SEQUENCE 450 AA; 50807 MW; E9717367 CRC32;

Query Match
Best Local Similarity 34.7%; Score 69; DB 13; Length 450;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 329 KVSDFGLTKKASAIQDTSKLVK 352
QY 1 KSSAYSQMGATAIKQVKLFKK 24

RESULT 8 PRELIMINARY; PRT; 1170 AA.
ID O48573;
AC O48573;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PUTATIVE DISEASE RESISTANCE PROTEIN.
GN T19K24.2; IS THALIANA (MOUSE-EAR CRESS).
OS ARABIDOPSIS THALIANA; VIRIDIPHYTES; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC EUPHYLOPHYTES; SPERMATOPHYTES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AC002342; G2660663;
SQ SEQUENCE 1170 AA; 133397 MW; 9B32952B CRC32;

Query Match
Best Local Similarity 34.7%; Score 69; DB 10; Length 1170;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 372 QIETLFGKWNW 382

```

```

QY 16 QVKKLFKKWGW 26

RESULT 9 PRELIMINARY; PRT; 157 AA.
ID O58919;
AC O58919;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE 157AA LONG HYPOTHETICAL PROTEIN.
GN PH177;
OS PYROCoccus HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCoccus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RL "Complete Sequence and Gene Organization of the Genome of a
Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA RES. 5:55-76(1998).
DR EMBL: AP000005; D1031220;
SQ SEQUENCE 157 AA; 18430 MW; 3A304921 CRC32;

Query Match
Best Local Similarity 34.2%; Score 68; DB 1; Length 157;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 125 SLEINKLTREEREELYKKWS 144
QY 6 SLQMGATAIKQVKLFKKW 25

RESULT 10 PRELIMINARY; PRT; 214 AA.
ID O49635;
AC O49635;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 23.1 KD PROTEIN.
GN MLC1581.27 OR B1177_C1_140.
OS MYCOBACTERIUM LEPRAE.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA OLIVER K., HARRIS D.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA PARKILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93186700.
RA EIGLMER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae."
RL MOL. MICROBIOL. 7:197-206(1993).
RN [4]
RP SEQUENCE OF 137-214 FROM N.A.
RA ROBISON K., SMITH D.R.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Z96801; E321707;
DR EMBL: U00011; G466808;
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE; SIGNAL.
SQ SEQUENCE 214 AA; 23114 MW; F162EC6B CRC32;

```

Query Match
Best Local Similarity 34.2%; Score 68; DB 2; Length 214;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 91 FPKHFGHVALVERLNRWG 111
QY 5 YSLQMGATAIKQVKLFRKW 25

RESULT 11
ID Q38150 PRELIMINARY; PRT; 256 AA.
AC Q38150;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DE PHAGE SP1 DNA SEQUENCE CODING FOR PRODUCTS REQUIRED FOR REPLICATION
DE INITIATION.
GN 38.
OS BACTERIOPHAGE SP1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE;
OC LAMEDA PHAGE GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94172631.
RA PEDRE X., WEISE F., CHAI S., LUEDER G., ALONSO J.C.;
RT "Analysis of cis and trans acting elements required for the
RT initiation of DNA replication in the Bacillus subtilis bacteriophage
RL J. MOL. BIOL. 236:1324-1340(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA ALONSO J.C.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X67865; G439642; -.
DR EMBL; X97818; E244864; -.
SQ SEQUENCE 256 AA; 29998 MW; 16837A2C CRC32;

Query Match
Best Local Similarity 34.2%; Score 68; DB 9; Length 256;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 60 ITSIRLNMNRWG 72
QY 14 IKQVKLFRKW 26

RESULT 12
ID Q91567 PRELIMINARY; PRT; 51 AA.
AC Q91567;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DE PROTEIN-TYROSINE KINASE (G56) (FRAGMENT).
GN G56.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDEA; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-EMBRYO;
RX MEDLINE; 95383727.
RA BRANDLI A.W., KIRSCHNER M.W.;
RT "Molecular cloning of tyrosine kinases in the early Xenopus embryo;
RT identification of Eck-related genes expressed in cranial neural crest
RL cells of the second (hyoid) arch."
RL DEV. DYN. 203:119-140(1995).
DR EMBL; U11729; G509579; -.
DR PFAM; PF00069; PKINASE; 1.
KW TYROSINE-PROTEIN KINASE.

FT NON_TER 1 1
SQ SEQUENCE 51 AA; 5577 MW; 84274507 CRC32;
Query Match
Best Local Similarity 33.7%; Score 67; DB 13; Length 51;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 13 KVSDFGLKEASAIQSSKLPVKW 36
QY 1 KSSAYSLOMGATAIKQVKLFRKW 24

RESULT 13
ID Q66072 PRELIMINARY; PRT; 225 AA.
AC Q66072;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DE VIRION PROTEIN (FRAGMENT).
OS CANINE HERPESVIRUS.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RA REMOND M., SHELDRICK P., LEBRETON F., NARDEUX P., FOULON T.;
RL J. VIROL. 77:37-48(1996).
DR EMBL; X90443; G1066229; -.
FT NON_TER 225 225
SQ SEQUENCE 225 AA; 25754 MW; E3E43F9B CRC32;

Query Match
Best Local Similarity 33.7%; Score 67; DB 14; Length 225;
Matches 8; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Db 111 AKVYGLQGEAVKQVADLFTTW 133
QY 2 SSAYSLOMGATAIKQVKLFRKW 24

RESULT 14
ID Q58328 PRELIMINARY; PRT; 457 AA.
AC Q58328;
DT 01-AUG-1998 (TREMELREL. 07, CREATED)
DT 01-AUG-1998 (TREMELREL. 07, LAST SEQUENCE UPDATE)
DE 457AA LONG HYPOTHETICAL PROTEIN.
GN PH0589.
OS PYROCOCCUS HORIKOSHII.
OC ARCHAEA; EURIARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL; AP000002; D1030621; -.
SQ SEQUENCE 457 AA; 52090 MW; CD2283F7 CRC32;

Query Match
Best Local Similarity 33.7%; Score 67; DB 1; Length 457;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 97 REVREYMRKMGW 108
QY 111

wed Sep 15 12:54:02 1999

15 KQVKKLFKKWG 26

RESULT	15				PRT:	646 AA.
ID	O37319				PRELIMINARY:	
AC	O37319:	(TREMUREL	05,	CREATED)		
DT	01-JAN-1998	(TREMUREL	05,	LAST SEQUENCE UPDATE)		
DT	01-JAN-1998	(TREMUREL	08,	LAST ANNOTATION UPDATE)		
DT	01-NOV-1998	(NUCLEOSIDE TRIPHOSPHATASE I	(EC 3.6.1.15)	(NUCLEOSIDE-TRIPHOSPHATASE)		
DE		(NUCLEOSIDE TRIPHOSPHATE PHOSPHOHYDROLASE).				
DE		(NUCLEOSIDE TRIPHOSPHATE PHOSPHODIOLASE).				
DE		(HELIOTITIS ARMERIA ENTOMOPOXIVIRUS (HAEPV)).				
OC		VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; ENTOMOPOXVIRINAE.				
RN	[1]	SEQUENCE FROM N.A.				
RP		STRAIN=CLONAL ISOLATE WT#2/011293;				
RC		SRIKANTHA A., OSBORNE R.J., DALL D.J.:				
RA		'SRIKANTHA A.', 0:0-0(1997).				
RL		-!- CATALYTIC ACTIVITY: NTP + H ₂ O = NDP + PHOSPHATE.				
CC		EMBL: AF019224; G2460172;				
DR		PFAM: PF00176; SNF2_N: 1				
DR		PFAM: PF00271; helicase_C: 1.				
KW		HYDROLASE.				
DQ		SEQUENCE	646 AA;	75763 MW;	Ic626ECF	CRC32;
SQ						

```

Query Match          33.7%; Score 87; 55
Best Local Similarity 54.5%; Pred. No. 3.26e+00;
Matches             6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

Db 75 IKHIKELFPRW 85
11:11:1
QY 14 IKQVKKLFKKW 24

Search completed: Wed Sep 15 10:39:47 1999
Job time : 25 secs.

 W P S R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:38:16 1999; MasPar time 262.06 Seconds
 Tabular output not generated. 1205.840 Million cell updates/sec

Title: >US-09-068-507A-2
 Description: (1-114) from US09068507A.seq
 Perfect Score: 114
 N.A. Sequence: 1 ATGATGATATTTAAACT.....AAATTTTACCCATCGTTAA 114
 Comp: TACTACTATAAATTTTGA.....TTTAAATGGGTAGCAAT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb158
 1:em_bal 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
 7:em_in 8:em_om 9:em_ov 10:em_ov 11:em_pat 12:em_ph
 13:em_pl 14:em_ro 15:em_sts 16:em_v1
 Database: genbank111
 17:gb_bal 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
 22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
 28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
 33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 9.118; Variance 6.864; scale 1.328

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	114	100.0	5873	17	Lactobacillus sake ind	7.64e-44
2	114	100.0	7597	17	LSSAKACNP	7.64e-44
3	31	27.2	10772	21	L.sake spIA, sppA, spp	3.81e-02
4	29	25.4	216021	31	Drosophila melanogaste	2.65e-01
5	29	25.4	216021	31	Homo sapiens Chromosom	2.65e-01
6	27	23.7	7218	25	Homo sapiens Chromosom	1.76e+00
7	26	22.8	1056	23	Sequence 14 from paten	4.43e+00
8	26	22.8	1056	23	Mustela vison GT dinuc	4.43e+00
9	26	22.8	7430	28	Campotheca acuminata	4.43e+00
10	26	22.8	10772	21	Drosophila melanogaste	4.43e+00
11	26	22.8	31933	22	Caenorhabditis elegans	4.43e+00
12	26	22.8	38856	21	Caenorhabditis elegans	4.43e+00
13	26	22.8	44996	21	Caenorhabditis elegans	4.43e+00
					Homo sapiens chromosom	4.43e+00

c	14	22.8	158078	31	AC005094	Homo sapiens BAC clone	4.43e+00
c	15	22.8	198526	20	AC006280	Plasmodium falciparum	4.43e+00
c	16	21.9	1891	21	DDU73685	Dictyostelium discoide	1.10e+01
c	17	21.9	3108	27	LELDHIGN	Lycopersicon esculentu	1.10e+01
c	18	21.9	6993	29	AB006625	Homo sapiens mRNA for	1.10e+01
c	19	21.9	20962	21	CEY48E1C	Caenorhabditis elegans	1.10e+01
c	20	21.9	28220	21	CEM176	Caenorhabditis elegans	1.10e+01
c	21	21.9	38225	21	CEY13G11	Caenorhabditis elegans	1.10e+01
c	22	21.9	73210	19	AC005653	Drosophila melanogaste	1.10e+01
c	23	21.9	80353	20	AC004710	Plasmodium falciparum	1.10e+01
c	24	21.9	178367	31	AC006115	Homo sapiens chromosom	1.10e+01
c	25	21.9	191908	19	CEY5189	Caenorhabditis elegans	1.10e+01
c	26	21.9	291821	19	CEY48E1	Caenorhabditis elegans	1.10e+01
c	27	21.1	868	28	AF006509	Endogone pisiformis 18	2.69e+01
c	28	21.1	871	28	AF006510	Endogone pisiformis 18	2.69e+01
c	29	21.1	878	28	AF006511	Endogone pisiformis 18	2.69e+01
c	30	21.1	25359	21	CEY02D10	Caenorhabditis elegans	2.69e+01
c	31	21.1	30363	21	CEY02D10	Caenorhabditis elegans	2.69e+01
c	32	21.1	39356	21	CEY32A11	Caenorhabditis elegans	2.69e+01
c	33	21.1	42696	21	CELC45G7	Caenorhabditis elegans	2.69e+01
c	34	21.1	48338	19	CEY02D9	Caenorhabditis elegans	2.69e+01
c	35	21.1	58663	22	AC004300	Drosophila melanogaste	2.69e+01
c	36	21.1	74371	31	AC005369	Homo sapiens chromosom	2.69e+01
c	37	21.1	82966	19	CEAA2	Caenorhabditis elegans	2.69e+01
c	38	21.1	120011	19	HS124K22	Human DNA sequence ***	2.69e+01
c	39	21.1	137678	29	HS879K22	Human DNA sequence fro	2.69e+01
c	40	21.1	182152	19	CEY67A10	Caenorhabditis elegans	2.69e+01
c	41	21.1	183422	19	AC006905	Caenorhabditis elegans	2.69e+01
c	42	21.1	195349	19	AC006705	Caenorhabditis elegans	2.69e+01
c	43	21.1	228418	20	AC005505	Plasmodium falciparum	2.69e+01
c	44	21.1	257817	19	AC006909	Caenorhabditis elegans	2.69e+01
c	45	21.1	299308	19	AC006898	Caenorhabditis elegans	2.69e+01

ALIGNMENTS

RESULT	1	AF002276	5873 bp	DNA	BCT	01-JAN-1998
LOCUS						
DEFINITION						
ACCESSION						
NID						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						


```

REMARK (sites)
REFERENCE 4 (bases 1 to 7597)
AUTHORS Huehne, K.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1995) Huehne K., Federal Centre for Meat
Research, Microbiology and Toxicology, E.-C.-Baumann-Strasse 20,
Kulmbach, Bavaria, FRG, D-95326
FEATURES
    source
        1..7597
            /organism="Lactobacillus sakei"
            /strain="Lb674"
            /db_xref="taxon:1599"
            412..417
            /-35_signal
            439..444
            /-10_signal
            557..562
            RBS
            572..1918
            gene
            /gene="sppK"
            572..1918
            CDS
            /gene="sppK"
            /function="putative histidine protein kinase"
            /citation=[1]
            /codon_start=1
            /transl_table=11
            /product="SppK"
            /protein_id="CAA88426.1"
            /db_xref="PID:g695616"
            /db_xref="GI:695616"
            /db_xref="SPTREMBL:Q48870"
            /translation="MLVTVSVSLMNFVAILLFLLRYIQRKIKFKRIILDLIAI
            IFSILYIFISDASLLVWMLRWLGHWHQKQENKIKTDTANGLIILVIQILLVAVGTI
            ISQTSIIISDFSEQNILNNSATDITLLGIFFAVFDGLFFILLKRNKRELQHLNQEI
            IEFLEQYIFIFILFIVIEIILAVNLQGVATILITIIIFCVLIGMTQVQMLF
            LKAYSIOEANDOLVRNQOQDYLVNIEQOYTTLRRFKHDYQNIILSLSEFAKQDQO
            QKAYVOELAAQRPQIOEIOGAVIAQLDYLNKNDPIRGLVIOKFLAAKQAGVTLKFMT
            EPIELANLITVIRIIGLLDNAIEQAQVETQDLYSCAFLOSDGLIETIENTASQV
            KNLQAFSELGTSGAGRGIGLANVQDLIAQKLNLEFQIENKRLKRLQILMITEE"
            1906..1910
            RBS
            /gene="sppK"
            1920..2666
            gene
            /gene="sppK"
            1920..2666
            CDS
            /gene="sppK"
            /function="putative response regulator protein"
            /note="ttg start codon"
            /citation=[1]
            /codon_start=1
            /transl_table=11
            /product="SppR"
            /protein_id="CAA88427.1"
            /db_xref="PID:g695617"
            /db_xref="GI:695617"
            /db_xref="SPTREMBL:Q48871"
            /translation="MYVYLLDDLQQAIVQIILANTIMINEFAMTLCAASDTEEL
            LAIKQQRGLFUDMEIDNRQGLVATKIKQMPFAQIVFITTHEELTLTLERK
            IAPDYLIKQOTMAEIKRQLDLDLLAEKQNEAAHYRENLFYKIGPRFFSPLEKY
            VLYTERENGHINLLATRKRVTFPGNVLAEQAQYPMLEFRCDKSYLVNLSNIANYDSK
            TRSLKFDGSGAKVSFRKSLVAKLQMM"
            2670..2706
            terminator
            2712..2717
            -35_signal
            2736..2741
            -10_signal
            2798..2802
            RBS
            2813..2998
            gene
            /gene="sppA"
            2813..2998
            CDS
            /gene="sppA"
            /standard_name="sakacin P preprotein"
            /function="bacteriocin preprotein"
            /citation=[1]
            /citation=[2]
            /citation=[3]
            /codon_start=1
            /transl_table=11
            /product="SppA"

/protein_id="CAA88428.1"
/db_xref="PID:g695618"
/db_xref="GI:695618"
/db_xref="SPTREMBL:Q57121"
/translation="MEXFIELSLKEVTAITGKYYGNVHCGHCKHSCIVDMWCTAIGNIG
NNAANWAIGGNAGWNK"
2813..2866
/gene="sppA"
/function="signal peptide"
2867..2995
/gene="sppA"
/function="bacteriocin"
/evidence=experimental
/product="sakacin P"
3079..3084
RBS
3092..3388
gene
/gene="sppA"
3092..3388
CDS
/gene="sppA"
/function="putative protein conferring immunity to sakacin
P"
/citation=[1]
/citation=[3]
/codon_start=1
/transl_table=11
/product="SpiA"
/protein_id="CAA88429.1"
/db_xref="PID:g695619"
/db_xref="GI:695619"
/db_xref="SPTREMBL:Q57161"
/translation="MKILKWSGKDRGERANDIIGQLLDLHDPKNEHLEAILNY
ONEIRKSSVPFILSRMNIANTIRDRILITDFOEDKLLTALSNIYGY"
3544..3594
terminator
3654..3659
-35_signal
3675..3680
-10_signal
3728..3732
RBS
3739..5895
gene
/gene="sppT"
3739..5895
CDS
/gene="sppT"
/function="putative ATP-dependent translocator protein"
/citation=[1]
/codon_start=1
/transl_table=11
/product="SppT"
/protein_id="CAA88430.1"
/db_xref="PID:g695620"
/db_xref="GI:695620"
/db_xref="SPTREMBL:Q48872"
/translation="MIDRKIYIPQVDEMDCGVAALAMILKHYGSSASLAYLRNEAKTD
LEGTTALGLVKTAEENLGPETKAIQADMSLFEVADLPFPFIAHLKNGELLHYVVLGI
KDAIILADPDPTVGIKMTOSQKSEWTVALTAPKPSQPVKQPKASLFSFVPT
LMQKRLVINITLAALLTIISLGSYFLQVVDITYIPNNMQSILAIYVAIGLIIYGF
LSIYTAQNFLLAVLQGLSIEILSYLRHVFLPMSFPATKRTGEIISRFENDANKII
DALASAIISIDITVVYIMGLIIQNTLFWLTAVIPIYIAVITLFAKPFKLNQ
KAMESNVLNIAIEDLHGIEIKALTSERYOKIDTAFDYILKSFYVYKADTLOQ
ATKLVQLGNVYLVGAILVTNLSIGOLMTYNALLAVFNPLONIINLOTKLOS
AVANRLNVEYLVDSSEKTERPKSESQLYGPIELQNVYQYGVQGRVLSISLAKIE
PNEKTIYVMGSGSGKSTLVKLLVNFETEGKLFNQDQDTIDKRLURQVNYIAQT
PYIFSGSDNLKLNRPGITQADIMHACELAEIKNDIEKMPLOPETILDDGNTLSG
GOKRLTARALLSPARVLIFFDESTGLDVLVDHLMALERTIIFIAHRLAIA
KKTNNILVLENGSVVETGTGTHALVDVTQGRYQLLNH"
5897..5901
RBS
5910..7286
gene
/gene="sppE"
5910..7286
CDS
/gene="sppE"
/function="putative accessory protein for sakacin P
export"
/citation=[1]
/codon_start=1
/transl_table=11
/product="SppE"

```

		/protein_id="CAA88431.1"				
		/db_xref="PID:g695621"				
		/db_xref="GI:695621"				
		/translation="MDKEFLSESEFYSAEKNEFTLLIMPIAVLLCLVCFSEFGKRE				
		ITIEGQGLTINKQIPILQASTNSVLYKQNLKQKGFYKKGQTLVYQNTKNQKRL				
		EKKQNLQITSLTETLKASIRSNQDFENNDFYRDLGLRGYDQDQRYLYENQYQ				
		DKAATSTTKQTLTAKTIRNQTNDLAYQALYQIKTNKYVSTGYNYLYENQYQ				
		TKVMSKDNQSGSLGKILATTOQDLSLQGSIAANAOKVSIKEFDSDSRFNNTN				
		EKLALLEAQOTQSAEOLVAKQSLVEIOTTLKQITSDSKYEVVAKPKDGLVHDDHY				
		OGIKYTSAGTSMARIYPLVDQKRLKIEALIPVDDISSVYKIGORLRLKIRNVPRII				
		TEGKINQSVSPFVINGSYIIIGLATISNNNRKLLHYGWTGKIALITGKTTFFNY				
		KDKLNRN"				
BASE COUNT	2636 a	1102 c	1482 g	2377 t		
ORIGIN						
	Query Match	100.0%	Score 114;	DB 17;	Length 7597;	
	Best Local Similarity	100.0%	Pred. No. 7.64e-44;			
	Matches 114;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
Db	440	ATGATGATATTTAAAAAAGCTTTTCAGAAAAAGAAATTCGAAAAATAAACGGTGGTATGGCA	499			
QY	1	ATGATGATATTTAAAAAAGCTTTTCAGAAAAAGAAATTCGAAAAATAAACGGTGGTATGGCA	60			
Db	500	GGAAATAGTCTCAATTTTATTCATAGATTAACAAAATTTTACCCTCGTTAA	553			
QY	61	GGAAATAGTCTCAATTTTATTCATAGATTAACAAAATTTTACCCTCGTTAA	114			
RESULT	3					
LOCUS	AF012089	10772 bp	DNA	INV	05-AUG-1997	
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete cds, and phenylalanyl tRNA synthetase gene, partial cds.					
ACCESSION	AF012089					
NID	92305220					
VERSION	AF012089.1	GI:2305220				
KEYWORDS	fruit fly.					
SOURCE	Drosophila melanogaster					
- ORGANISM	Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 4546 to 4553)					
AUTHORS	Gray, Y.H., Tanaka, M.M. and Sved, J.A.					
TITLE	P-element-induced recombination in Drosophila melanogaster: hybrid element insertion					
JOURNAL	Genetics 144 (4), 1601-1610 (1996)					
MEDLINE	97132596					
REFERENCE	2 (bases 1 to 10772)					
AUTHORS	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.					
TITLE	Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects					
JOURNAL	unpublished					
REFERENCE	3 (bases 1 to 10772)					
AUTHORS	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.					
TITLE	Direct Submission					
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia					
FEATURES	Location/Qualifiers					
source	1..10772					
	/organism="Drosophila melanogaster"					
	/db_xref="taxon:7227"					
mRNA	join(872..1000,2310..2426,6476..6690,6751..7707)					
	/gene="Cp1"					
gene	872..7707					
	/gene="Cp1"					
exon	872..1000					
	/gene="Cp1"					
	/number=1					
intron	1001..2309					
	/gene="Cp1"					
exon	2310..2426					
/gene="Cp1"						
/number=2						
CDS	join(2328..2426,6476..6690,6751..7462)					
/gene="Cp1"						
/codon_start=1						
/product="cysteine proteinase-1"						
/protein_id="AAB65749.1"						
/db_xref="PID:g2305221"						
/db_xref="GI:2305221"						
/translation="MRTAVLLPLALLAQAQVSPADVVVEWHTEKLEHRKNYQDET						
EERPLKIFENKHKIAKHNRQAEKGVSKLAVNKYADLLHHEFRLMNGNYTLHR						
QKRAQSGVLYSEQLVDCSTKYKNGCNGGLMDNAPRIKONGGIDTEKSYPEYAE						
HFRKSGVLSQNLVDCSTKYKNGCNGGLMDNAPRIKONGGIDTEKSYPEYAE						
DSCHFNGTGVATDRGFTDIPQGEKKMAEAVATGVPVVAIDASHESQFQSEGVYN						
EPOCDQNLDHGVLYVVGFTDESGEDYWLWNKSWGTTGDKGFIKMLRKNKQCGIAS						
ASSYPLV"						
intron	2427..6475					
	/gene="Cp1"					
misc_feature	4546..4553					
	/gene="Cp1"					
	/note="insertion site of P(Casper)(50C)"					
exon	6476..6690					
	/citation=[1]					
	/gene="Cp1"					
	/number=3					
intron	6691..6750					
	/gene="Cp1"					
exon	6751..7707					
	/gene="Cp1"					
	/codon_start=1					
	/product="phenylalanyl tRNA synthetase"					
	/protein_id="AAB65750.1"					
	/db_xref="PID:g2305222"					
	/db_xref="GI:2305222"					
	/translation="MLTLRVQARHWLSTRCLASAPAKSPSPPOLEVSGSTYIA					
TDGNTVTPKILSYGANKHLOTDPHLSIRQIRVINYFYGAYRNQGNPLFSYDQMN						
PVTVQQNFQNLIPADHVSQRKDCYYINQQLHRAHTAHQVELISGLGNFLVVG						
EYVRDEIDSTHYFVHQADAVRLTKDLFERNPGLSEFTWSTGLADPLILPH						
SSWTKPNSPATRAVKLMEMKHVLVGLTKDLFGPRIKRVVDYFPTQPSWELEI						
YFKDWLEVLGCGIMRHEILQSRGSHQSIGYAFGVLERLAWLEPIDIRLFWNSD						
GFLSQFSEKDLNLPKIPISHYPQTDNLDSFLPQDIEVDAGFSPNDYDLRVSVAG						
DMVEQISLVDFKFKPKTKGSSVCFRIVYRHMERTLTOAEVNEIHKQIASVDSFNVQ						
IX"						
BASE COUNT	2929 a	2357 c	2282 g	3046 t	158 others	
ORIGIN						
	Query Match	27.2%	Score 31;	DB 21;	Length 10772;	
	Best Local Similarity	16.7%	Pred. No. 3.81e-02;			
	Matches 17;	Conservative	52;	Mismatches 32;	Indels 1;	Gaps 1;
Db	1729	WTSRTTTTSAMWMTWTSTWTKYMWAYAWKMWKMTTCTWMAWASWARKWKTSA	1787			
QY	1	ATGATGATATTTAAAAAAGCTTTTCAGAAAAAGAAATTCGAAAAATAAACGGTGGTATGGCA	60			
Db	1788	SAWRKMWKWAYRAMKKTWMAAWKRWKAAWMTWRWYMTT	1829			
QY	61	GGAAAT				


```

RESULT 9
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cpl) gene, complete
cds, and phenylalanine tRNA synthetase gene, partial cds.
ACCESSION AF012089
NID Q2305220
VERSION AF012089.1 GI:2305220
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Echinodermata; Metazoa; Arthropoda;
Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 4546 to 4553)
AUTHORS Gray, Y.H., Tanaka, M.M. and Sved, J.A.
TITLE P-element-induced recombination in Drosophila melanogaster: hybrid
element insertion
JOURNAL Genetics 144 (4), 1601-1610 (1996)
MEDLINE 97132596
REFERENCE 2 (bases 1 to 10772)
AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
TITLE Structure of the cysteine proteinase (Cpl) gene of Drosophila
melanogaster and associated mutational effects
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 10772)
AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University
of Sydney, Biology A12, Sydney University, NSW 2006, Australia
FEATURES
Location/Qualifiers
1..10772
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
join(872..1000,2310..2426,6476..6690,6751..7707)
/gene="Cpl"
/product="cysteine protease"
872..7707
/gene="Cpl"
exon 872..1000
/gene="Cpl"
intron 1001..2309
/gene="Cpl"
exon 2310..2426
/gene="Cpl"
number=2
join(2328..2426,6476..6690,6751..7462)
/gene="Cpl"
CDS /codon_start=1
/product="cysteine proteinase-1"
/protein_id="AAB65749.1"
/db_xref="PID:g2305221"
/db_xref="GI:2305221"
/translat="MRTAVLLPALLAVAAQVSPADVVVEWHFLEHKNYQDET
EERFLKFNENKHTKAKNOREAEKVSFKLAVKAYADLLHFERQLMNGFNTHLK
OLRADESFGVTFISPAHVTLPKSVDRITKGAATVAKDQGGHGCSCAFSGALEGQ
HFRKSVLSQNLVDCSTKGNKNGCNGMLDNFRIKNGDITKSYFEAID
DSCHFNGTGTADTDFDIPOGDEKMAEAVATGVPVSAVDASHESFQFSEGVIN
EPQDAQNLHDHGLVYVFGTDESGDYLVKNSWGTGDKGFIKMLRNKNQCGIAS
ASSYPLV"
intron 2427..6475
/gene="Cpl"
misc_feature 4546..4553
/gene="Cpl"
/note="insertion site of P{Casper}{50C}."
/citation=[1]
6476..6690
/gene="Cpl"
number=3
6691..6750
/gene="Cpl"
intron 6751..7707
/gene="Cpl"
exon

```

```

/number=4
join(<8110..9300,9370..>9532)
/product="phenylalanine tRNA synthetase"
join(8110..9300,9370..>9532)
/note="potential orf"
/codon_start=1
/product="phenylalanine tRNA synthetase"
/protein_id="AAB65750.1"
/db_xref="PID:g2305222"
/db_xref="GI:2305222"
/translat="MLLTLRVQGARHMLKSTRCLASSAAPKSPSPQLEVGSGSTYA
TDCMTNVPKILSYGANKHLQTDHPLSIIRQIVNVEFYGAYRNQGNPLFSYVDQMN
PVYTVQNFNDLLIPADHVSQKSCYIINQHLRLRAHTAHQVELISGLDNLFLVGG
EYVREDESDTHYFVPHQADAVRLVKDLKLFERNPGLFEETWGTGLPKLLPHP
SWTAPNPAIRNPAVKLMEHMKHVLVGLTKDLFGPRIKRYWDTYFFPTQPSWELEI
YFKDNWLVGCGIMRHEILQRSVGHQSIGYAFGVGLERLAWLFDIDIRLFSNDS
GFLSQFSEKDLHNLFPKYPI SHYPOCTNDLSFWLPODIEVDAGSPDNFDLVRVAG
DMVEQISLVDFKFKPKTKCKSSVCFRIVYRHMTLTQAEVNEIHKQIASASVDSENVQ
IX"
BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
ORIGIN
Query Match 22.8%; Score 26; DB 21; Length 10772;
Best Local Similarity 19.7%; Pred. No. 4.43e+00;
Matches 13; Conservative 33; Mismatches 20; Indels 0; Gaps 0;
Db 1693 RWYWAATTTMKWMTTWKWKMMKTYRTWKKWYWSRTTTTSSAMMMWYTWSTWYK 1752
Cp 71 GAACTATTTCCTGCCAATACCCGTTATTTTTCGAATCTTTTCTGAAAGTTTATA 12
Db 1753 WAYAWM 1758
Cp 11 AATATC 6
RESULT 10
LOCUS CELY50F7A 31933 bp DNA INV 06-FEB-1999
DEFINITION Caenorhabditis elegans cosmid Y50F7A.
ACCESSION AF125457
NID 94226117
VERSION AF125457.1 GI:4226117
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittidae;
Rhabdittina; Rhabdittidae; Rhabdittidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 31933)
AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Shownkeen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Woldman, P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 31933)
AUTHORS Strong, C., Tin-Wollam, A. and Harper, M.
TITLE The sequence of C. elegans cosmid Y50F7A
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 31933)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center

```


dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 This sequence is the entire insert of clone C36B1. The true right end of clone C36B1 is at 17898 in this sequence. The start of this sequence (.110) overlaps with the end of sequence 281079. The end of this sequence (44993..44996) overlaps with the start of sequence 236047.

FEATURES

Source

Location/Qualifiers

1. .44996
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="I"
 /clone="C36B1"

gene

/gene="C36B1.1"

CDS

complement(506..5930)
 complement(join(506..619,868..1025,1081..1154,1205..1302,1357..1466,5816..5930))
 /gene="C36B1.1"
 /note="CDNA EST EMBL:D34750 comes from this gene"
 /codon_start=1
 /protein_id="CAB02278.1"
 /db_xref="PID:el344394"
 /db_xref="PID:g3874785"
 /db_xref="GI:3874785"
 /db_xref="SPTREMBL:Q93336"

gene

/translation="MLPRHIGGLYFCALYLLRASAQPTFFALNALPGDSETEV
 DLPADKASLNADARVAKGIESPAIGLQGVETAPVRLYLPNFFKQFSLIATI
 KLMDKRGVLDFAVDASADIGLIEPAGTKQISLIVRSVAIVSLFVDFDSOQW
 TOPALEVVDQTVTFEYKCRFASRQVSLPDSFDEAKLYIASAGPIIDNGFEVSLF
 LILL"
 complement(7075..10843)

CDS

/gene="C36B1.2"
 complement(join(7075..7179,7290..7366,7941..8189,8242..8411,8742..9000,9134..9328,10759..10843))
 /gene="C36B1.2"
 /note="similar to fibronectin type-III domain"
 /codon_start=1
 /protein_id="CAB02267.1"
 /db_xref="PID:el344386"
 /db_xref="PID:g3874774"
 /db_xref="GI:3874774"
 /db_xref="SPTREMBL:Q93337"

gene

/translation="MIETGSQYLLFLFTCSLNSSTALQHEDRVNPAPQNVRIKTO
 STSATLWADPPTVLIGITVEIGESISQKILLEGPDSTSTFVIRLSPNTNYVFA
 VSAVNAEGEDTKVMAAKTRPESGQTEKLPPTSVRARIDERSAAGSAFVSWDDP
 NPSSSENSIDTKQYVINYIYESDTQOKVRSNAKAVRLTLGIPKEYEYAVKYVA
 GDGRSPWSIRDLVPEKTVSKFDFWCLNDTEMCSIHSSPHWKLCEKHDTYTOR
 DAGACPRVOYPSPAHLTPALNLPDAQRLCLYFALLNFPQMKVEIFRDGDMAN
 KQIVKTRMANVSVSLKWDGQPPRVIVHEMDVLSGCSERSGK"
 complement(13903..15051)

CDS

/gene="C36B1.4"
 complement(join(13903..14016,14097..14393,14440..14691,14933..15051))
 /gene="C36B1.4"

/note="similar to proteasome A-type submit: CDNA EST EMBL:M79843 comes from this gene; CDNA EST EMBL:D74896 comes from this gene; CDNA EST yk361a1.3 comes from this gene; CDNA EST yk361a1.5 comes from this gene; CDNA EST yk412d8.3 comes from this gene; CDNA EST yk412d8.5 comes from this gene"

/codon_start=1

/protein_id="CAB02269.1"

/db_xref="PID:el344385"

/db_xref="PID:g3874776"

/db_xref="GI:3874776"

/db_xref="SPTREMBL:Q95005"

/translation="MNRVDRALIFSPDGHFOVEAQAQVKKGSTAVGVKDCIVI
 GVKKSPALQDRTIRKIMIDHVMYLAFAGISADARVLDVDRARICQSKYKLTLED
 VTVAYISRYTANTKQRTSPGRPPFGISMLIGFDHGDGIFRLFKIEFSPGAYEYVAN

gene

/gene="C36B1.3"

CDS

join(15333..15455,15510..15644,15692..15817,15970..16217,16267..16355,16401..16524,16659..16865,16912..16963,17223..17327)
 /gene="C36B1.3"
 /note="similar to DNA directed RNA polymerase II; CDNA EST EMBL:T00665 comes from this gene; CDNA EST EMBL:T01799 comes from this gene; CDNA EST EMBL:T01866 comes from this gene"
 /codon_start=1
 /protein_id="CAB02268.1"
 /db_xref="PID:el344384"
 /db_xref="PID:g3874775"
 /db_xref="GI:3874775"
 /db_xref="SPTREMBL:Q93338"

/translation="MPYANQNIETVTELTNDIKFVLWDTLSVANSLLRVFMAEYPT
 LAIDWQIETNTSVLHDEFIAHRLGLIFISDYHVKKQYTRDCAEFCECSIPFI
 LQMKCKDEATLAVTTEHLMPHNLTOTVRPACGKALRERGSTRDFEHNREELIVKLR
 KQELNLKAYVKKGSGEKHAKWNPCTGVAFYDVPDNLARHTIYPVNEEPRSDHSLP
 EDSTEKEAFFEPDNAPNFWFSIEGTGALPAQRIVTMGIGILKRLLELNMAISNELQ
 AHAQQHILIFHEVKKHTSYNTAFFSONMVCGCEKRLKRVGVDPYRNKKVNRNADGS
 GKPTVTKNRLIGVKKATIGVAGCKLCKMLIHQPGSHYCVSTCAYQKGCAMCGKKIQ
 NTKGLRQSTT"
 18732..19246

gene

/gene="C36B1.6"

CDS

join(18732..18934,18987..19033,19065..19246)
 /gene="C36B1.6"

/codon_start=1

/protein_id="CAB02271.1"

/db_xref="PID:el344387"

/db_xref="PID:g3874778"

/db_xref="GI:3874778"

/db_xref="SPTREMBL:Q93340"

/translation="MPGKECVILFVSVLTVLHNFKMFNFSEDKITCDTHFRLLHS
 VKTQYLEYFPNNGNINCLYCTCLDFSSNCSIDVFNISNPIMWICIASLAWIQ
 VMTESFPCKVDRSVKMKMECSIDCCSIFSVLLNIAIARMVMEGG"
 complement(20092..22271)

gene

/gene="C36B1.5"

complement(join(20092..20199,20247..20454,20533..20763,21060..21177,21227..21394,21441..21720,21763..21900,21967..22119,22185..22271))
 /gene="C36B1.5"

/note="similar to U4/U6 small nuclear ribonucleoprotein"

/codon_start=1

/protein_id="CAB02270.1"

/db_xref="PID:el344386"

/db_xref="PID:g3874777"

/db_xref="GI:3874777"

/db_xref="SPTREMBL:Q93339"

/translation="NAENGNAFVAPPPROFGSLANAESVAILNAOQNHGPTVSLER
 MEVSNQADSRHDAEMFAEDRRRARTLTPTDDVQVQLKRLALNQICLFGDADLR
 RKRLALLTINSEDEIAVLTHTDEVNADKADSETVYHGRGPIELRMARVSIADFSR
 KAKRLUDKAREEAPRAHAKALQAQHWQVQINHUASQVADTRVAFVCEFSADSEH
 IYTAGWSQVAVYKREQCAQEKFTGHSSQACARHPGAFQNDYSSLNVSQSDYQ
 TVLLWSLSQESPIGEQHPQVQKVAHPNGHHLATACFDSTWRMYDTLTKKELLYQ
 EHKSTQVADVAFHPDGSVALTGCHDCYGRVMDRTGRCIMFLDGHKKEITHSVEMPNQ
 YEMITGSSNSMKVMDLRNRTYTMPAITSVYTVRADAAGQYLYSASFDCITLKMS
 TTGWQPLRLQGHDRILCLVDISPDQWMCSSAFDRTFKLWQSDY"
 22786..23487

/gene="C36B1.7"

join(22786..22920,22965..23095,23142..23294,23337..23487)

/gene="C36B1.7"

/note="similar to di-hydrofolate reductase; CDNA EST

EMBL:D72539 comes from this gene; CDNA EST EMBL:D75436

comes from this gene"

/codon_start=1

/protein_id="CAB02272.1"

/db_xref="PID:el344388"

/db_xref="PID:g3874779"

/db_xref="GI:3874779"

/db_xref="SWISS-PROT:Q93341"
 /translation="MRKNLIIVAMDAEGGIGKNGVLPWRIKKMQYFASVTKNVSOS
 KRAVLGRKQWESIVTRRPLAGRLNIVLSQLPKQSDYIVVNSLEAMKLLSEP
 PVDSEIETIWNIGGAEIYDLALRENLDVDEIHLTRFKNFADVHLKSLDFSKMEKVN
 AEVSSENSEIPEENGLKFECKWKVVENH"
 23636..28511
 /gene="C36B1.8"
 join(23636..23782,23832..24032,24475..24625,24685..24872,
 24971..25047,25100..25193,25241..25570,25668..25987,
 26036..26268,26315..26553,26602..26710,26756..26992,
 ...
 Note: remainder of annotations omitted.

Query Match 22.8% Score 26; DB 21; Length 44996;
 Best Local Similarity 88.2%; Pred. No. 4.43e+00;
 Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Db 36007 TTTATTCGAATTTTTCGAGAAAGTTTAAAA 36040
 Cp 43 TTTTTCGAATTTTTCGAGAAAGTTTAAAA 10

RESULT 13
 LOCUS AC002489 91638 bp DNA HTG 20-JAN-1998
 DEFINITION Homo sapiens chromosome X, clone 592, WORKING DRAFT SEQUENCE, 8
 unorderd pieces.
 ACCESSION AC002489
 NID 92795820
 VERSION AC002489.1 GI:2795820
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 91638)
 AUTHORS Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and
 Lander,B.

TITLE Homo sapiens chromosome X, clone 592
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 91638)
 AUTHORS Hawkins,T.L., Birren,B.W., Fasman,K.H., Nusbaum,C., Lander,B.S.,
 McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
 Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Gerlanger,K.,
 Haggis,B., Huang,J., Hui,L., Jacotot,L., Kirby,A., Lane,M.,
 Mackenzie,J., Marquis,N., McDermott,J., Molla,M., Morrow,J.,
 Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T., Olotu,A.,
 Peterson,K., Reeve,M.P., Roberts,D., Rollins,G., Stilwell,J.,
 Stone,C., Strickland,C., Sydney,K., Tang,L., Wilmer,F., Zemtseva,I.
 and Zody,M.
 Direct Submission
 Submitted (22-AUG-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 20, 1998 this sequence version replaced gi:2772532.
 The Staden databases, finishing information, and all
 chromatographic files used in the assembly of this clone are
 available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 8591: contig of 8591 bp in length
 * gap of unknown length
 * 8592 28835: contig of 20244 bp in length
 * gap of unknown length
 * 28836 40356: contig of 11521 bp in length

* 40357 58902: contig of 18546 bp in length
 * gap of unknown length
 * 58903 61812: contig of 2910 bp in length
 * gap of unknown length
 * 61813 71640: contig of 9828 bp in length
 * gap of unknown length
 * 71641 75199: contig of 3559 bp in length
 * gap of unknown length
 * 75200 91638: contig of 16439 bp in length.
 FEATURES
 Location/Qualifiers
 1..91638
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="592"
 /clone_lib="unknown"
 /map="X"
 /chromosome="X"

BASE COUNT 25406 a 19091 c 18992 g 28011 t 138 others
 ORIGIN

Query Match 22.8% Score 26; DB 19; Length 91638;
 Best Local Similarity 88.2%; Pred. No. 4.43e+00;
 Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Db 34433 AAAAACTTTCAGAAAAAGCATTTTAAAAAGATAA 34466
 QY 13 AAAAACTTTCAGAAAAAGCATTTTAAAAAGATAA 46

RESULT 14
 LOCUS AC005094 158078 bp DNA PRI 27-AUG-1998
 DEFINITION Homo sapiens BAC clone RG332B22 from 7p15-p21, complete sequence.
 ACCESSION AC005094
 NID 93478660
 VERSION AC005094.1 GI:3478660
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 158078)
 AUTHORS Edwards,J. and Ozersky,P.
 TITLE The sequence of Homo sapiens BAC clone RG332B22
 JOURNAL Unpublished (1998)
 REFERENCE 2 (bases 1 to 158078)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 Submitted (12-JUN-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 3 (bases 1 to 158078)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 Submitted (27-AUG-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Aug 27, 1998 this sequence version replaced gi:3212897.

COMMENT
 SUBMITTED BY: WUGSC
 Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,

such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone RG332B22 is from the first release of the human BAC library CITB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:18794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RG023M10, 200 bp overlap. Actual start of this clone is at base position 1 of RG332B22; actual end is at 21981 of RG023M10.

FEATURES

source	Location/Qualifiers
STS	1..158078 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7p15-p21" /clone_lib="CITB-978SK-B" /clone="RG332B22" complement(4183..4302) /db_xref="GI:454632"
repeat_region	5685..5885 /rpt_family="Alu"
repeat_region	6316..6774 /rpt_family="L1"
repeat_region	6376..7114 /rpt_family="L2"
repeat_region	7454..7659 /rpt_family="MIR"
repeat_region	7747..8012 /rpt_family="Alu"
repeat_region	8815..9124 /rpt_family="Alu"
repeat_region	9384..9440 /rpt_family="MIR"
repeat_region	10844..11019 /rpt_family="L2"
repeat_region	11795..12088 /rpt_family="L1"
repeat_region	12304..12600 /rpt_family="Alu"
repeat_region	12877..13179 /rpt_family="Alu"
repeat_region	14006..14179 /rpt_family="Alu"
repeat_region	14600..14678 /rpt_family="MER81"
repeat_region	15150..15521 /rpt_family="MaLR"
repeat_region	15530..15826 /rpt_family="Alu"
repeat_region	16749..17049 /rpt_family="Alu"
repeat_region	17053..17133 /rpt_family="MaLR"
repeat_region	17157..17290 /rpt_family="Alu"
repeat_region	17294..17591 /rpt_family="L1"
repeat_region	17617..17896 /rpt_family="MaLR"
repeat_region	19068..19369 /rpt_family="Alu"
repeat_region	19842..20012 /rpt_family="Alu"
repeat_region	20131..20434 /rpt_family="Alu"
repeat_region	21967..22200 /rpt_family="Alu"
repeat_region	23876..24007 /rpt_family="MIR"
repeat_region	24222..25020 /rpt_family="MER21_g"
repeat_region	25430..25805 /rpt_family="MaLR"
repeat_region	25806..25860 /rpt_family="MIR"
repeat_region	25991..26070 /rpt_family="MIR"
repeat_region	26083..26295 /rpt_family="MER1_type"
repeat_region	26317..26548 /rpt_family="MIR"
repeat_region	27048..27148 /rpt_family="Alu"
repeat_region	27169..27432 /rpt_family="Alu"
repeat_region	27585..27790 /rpt_family="L2"
repeat_region	27852..28204 /rpt_family="MER2_type"
repeat_region	28407..28808 /rpt_family="L1"
repeat_region	29146..29736 /rpt_family="MER2_type"
repeat_region	30340..31208 /rpt_family="MER2_type"
repeat_region	32747..32816 /rpt_family="L2"
repeat_region	33065..33407 /rpt_family="L2"
repeat_region	33470..33529 /rpt_family="L2"
repeat_region	34582..35314 /rpt_family="L1"
repeat_region	35993..36158 /rpt_family="L2"
repeat_region	36171..36218 /rpt_family="L1"
repeat_region	36219..36392 /rpt_family="Alu"
repeat_region	36393..37446 /rpt_family="L1"
repeat_region	37449..37744 /rpt_family="Alu"
repeat_region	37745..37983 /rpt_family="L1"
repeat_region	38043..38104 /rpt_family="L2"
repeat_region	38554..38728 /rpt_family="MER1_type"
repeat_region	38733..39032 /rpt_family="Alu"
repeat_region	39339..39424 /rpt_family="BC200"
repeat_region	39426..39706 /rpt_family="Alu"
repeat_region	40645..40760 /rpt_family="Alu"


```

repeat_region 41284..41544
/rpt_family="MIR"
repeat_region 41601..41886
/rpt_family="Alu"
repeat_region 42295..42689
/rpt_family="MaLR"
repeat_region 42881..43253
/rpt_family="MaLR"
repeat_region 44287..44415
/rpt_family="MIR"
repeat_region 44788..44898
/rpt_family="L1"

```

... Note: remainder of annotations omitted.

Query Match 22.8%; Score 26; DB 31; Length 158078;

Best Local Similarity 79.5%; Pred. No. 4.43e+00;

Matches 35; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 62456 TTTATTTTACAAATTTTCTTTTAAATTTTAAATAAT 62499

Cp 47 TTTATTTTGCATTTCTTCTGAAGCTTTTAAATATCAT 4

RESULT 15

```

LOCUS AC006280 198526 bp DNA HTG 05-MAR-1999
DEFINITION Plasmodium falciparum 3D7 chromosome 12 PFYAC724 genomic sequence,
WORKING DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION AC006280
NID 94337163
VERSION AC006280.2 GI:4337163
KEYWORDS HTG; HTGS-PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 198526)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B. and Davis, R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 198526)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA

```

COMMENT

On Mar 5, 1999 this sequence version replaced gi:4176339.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

```

1 7090: contig of 7090 bp in length
* 7091 7290: gap of unknown length
* 7291 17234: contig of 9944 bp in length
* 17235 17434: gap of unknown length
* 17435 29116: contig of 11682 bp in length
* 29117 29317: gap of unknown length
* 29317 47101: contig of 17785 bp in length
* 47101 47301: gap of unknown length
* 47301 66040: contig of 18739 bp in length
* 66040 66240: gap of unknown length
* 66240 88294: contig of 22054 bp in length
* 88294 88494: gap of unknown length
* 88494 113409: contig of 24915 bp in length
* 113410 113609: gap of unknown length
* 113610 148826: contig of 35217 bp in length
* 148827 149026: gap of unknown length
* 149027 198526: contig of 49500 bp in length.

```

FEATURES

source 1..198526

```

/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"

```

BASE COUNT 79790 a 19005 c 19692 g 78435 t 1604 others
 ORIGIN

Query Match 22.8%; Score 26; DB 20; Length 198526;
 Best Local Similarity 86.1%; Pred. No. 4.43e+00;
 Matches 31; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 65805 AAAAACCTTTTCAAAAAAATTCCTCAAAAAAATAC 65840
 QY 13 AAAAACCTTTTCAAAAAAATTCCTCAAAAAAATAC 48

Search completed: Wed Sep 15 10:47:20 1999
 Job time : 544 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:53:11 1999; MasPar time 51.25 Seconds
Tabular output not generated. 476.662 Million cell updates/sec

Title: >US-09-068-507A-2
Perfect Score: (1-114) from US09068507A.seq
N.A. Sequence: 114
Comp: 1 ATGATGATATTTAAATAACTAAATTTTACCATCGTTAA 114
TACTACTATTAATTTTGATTTAAATGGTAGCAATT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 7.049; Variance 6.074; scale 1.161

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description	Pred. No.
		Match						
1	114	100.0		114	34	T67141	IF gene for bacterioc	4.71e-41
2	30	26.3		91	9	Q51746	Oligonucleotide probe	8.34e-03
3	30	26.3		91	46	V44650	Mammalian DNA replica	8.34e-03
4	30	26.3		91	9	Q51746	Oligonucleotide probe	8.34e-03
5	29	25.4		204	1	N81164	Base substituted E.co	2.10e-02
6	29	25.4		204	1	N81164	Base substituted E.co	2.10e-02
7	28	24.6		501	3	N50024	Sequence encoding new	5.26e-02
8	27	23.7		91	46	V44650	Mammalian DNA replica	1.31e-01
9	26	22.8		501	3	N50023	Sequence encoding new	3.21e-01

c	10	26	22.8	501	3	N50026	Sequence encoding new	3.21e-01
	11	24	21.1	114	12	Q70467	Generic DNA sequence	1.89e+00
	12	24	21.1	498	3	N50034	Sequence encoding new	1.89e+00
	13	24	21.1	501	3	N50032	Sequence encoding new	1.89e+00
	14	24	21.1	501	3	N50031	Sequence encoding new	1.89e+00
	15	24	21.1	501	3	N50033	Sequence encoding new	1.89e+00
	16	24	21.1	501	3	N50029	Sequence encoding new	1.89e+00
	17	24	21.1	501	3	N50028	Sequence encoding new	1.89e+00
c	18	24	21.1	501	3	N50033	Sequence encoding new	1.89e+00
c	19	23	20.2	67	24	T14322	Primer used in the la	4.50e+00
c	20	23	20.2	114	12	Q70465	Generic DNA sequence	4.50e+00
c	21	23	20.2	114	12	Q70469	Generic DNA sequence	4.50e+00
c	22	23	20.2	114	12	Q70468	Generic DNA sequence	4.50e+00
c	23	23	20.2	431	19	T19913	Human gene signature	4.50e+00
	24	23	20.2	675	44	V11706	Carnobacteriocin Bm1	4.50e+00
	25	22	19.3	501	3	N50025	Sequence encoding new	1.06e+01
	26	22	19.3	501	3	N50030	Sequence encoding new	1.06e+01
	27	22	19.3	501	3	N50027	Sequence encoding new	1.06e+01
c	28	22	19.3	501	3	N50026	Sequence encoding new	1.06e+01
c	29	22	19.3	501	3	N50030	Sequence encoding new	1.06e+01
c	30	22	19.3	501	3	N50023	Sequence encoding new	1.06e+01
c	31	22	19.3	501	3	N50032	Sequence encoding new	1.06e+01
c	32	22	19.3	501	3	N50024	Sequence encoding new	1.06e+01
	33	22	19.3	1308	44	V35184	Thuja plicata sinol-lar	1.06e+01
	34	22	19.3	2183	1	N90700	Clal fragment contg.	1.06e+01
	35	22	19.3	2976	17	T04687	Black widow spider de	1.06e+01
c	36	22	19.3	3440	3	Q13967	Bce-4 gene with regul	1.06e+01
c	37	22	19.3	3440	40	V15268	Brassica campestris B	1.06e+01
	38	22	19.3	3706	17	T04688	Human REST protein de	1.06e+01
	39	21	18.4	1090	24	T41410	Human neuron restrict	2.46e+01
	40	21	18.4	2043	23	T36898	Human REST protein co	2.46e+01
	41	21	18.4	3291	24	T41405	Insert of pSRQ220 con	2.46e+01
	42	21	18.4	5595	4	Q25810	NC-contig derived fro	2.46e+01
	43	21	18.4	80240	59	V83940	Mycoplasma genitalium	2.46e+01
	44	21	18.4	580073	27	T58840		

ALIGNMENTS

RESULT 1
ID T67141 standard; DNA; 114 BP.
AC T67141;
DE 10-DEC-1997 (first entry)
DT IF gene for bacteriocin-inducing peptide.
KW IF gene; gene expression; promoter; lactic acid bacterium;
KW bacteriocin; vaccine; vector; ss.
OS Lactobacillus sake strain LTH 673.
FH Key Location/Qualifiers
FT signal_peptide 1..54 /*tag= a
FT mat_peptide 55..111 /*tag= b
FT /*product= mature IF peptide
FN WO9718316-A1.
PD 22-MAY-1997.
PF 13-NOV-1996; NO0266.
PR 13-NOV-1995; NO-004575.
PA (BRUR/) BRURBERG M B.
PA (EJJS/) EJJSINK V G H.
PA (NESI/) NES I F.
PI Brurberg MB, Ejjsink VGH, Nes IF;
DR P-PSDB; W17988.
PT Gene expression system providing regulated lactic acid bacteria
PT protein production uses new inducing peptide involved in
PT bacteriocin synthesis, useful in fermentation and as a drug delivery
PT system
PS Example 4; Fig 2; 39pp; English.
CC This DNA sequence comprises the IF gene of Lactobacillus sake
CC LTH673 that encodes a secreted peptide (see W17988) which
CC autoinduces the expression of the IF-K-R gene cluster involved in
CC sakacin C production. The invention relates to the discovery of a
CC new regulatory mechanism for gene expression in lactic acid

CC bacteria that includes previously unrecognized, strongly
CC regulatable promoter elements (see T67142-48). Expression of genes
CC under the control of such promoters is dependent upon expression of
CC the IF-K-R gene cluster or an analogue. Thus, the expression of a
CC gene under the control of the promoter element in cells that
CC contain the K and R genes can be induced at will by adding the
CC cognate, synthetic (harmless) inducing peptide. A recombinant
CC vector containing the gene expression system, and host cells
CC transformed with this vector, incorporated in the chromosome,
CC and/or having integrated into its chromosome a promoter linked to
CC an integrated gene encoding a specific protein are claimed.
CC Typical applications are in fermentations (e.g. where the gene for
CC an enzyme is regulated) and production of specific proteins, or
CC where the bacteria express a surface antigen, as vaccines.
SQ Sequence 114 BP; 50 A; 12 C; 16 G; 36 T;
Query Match 100.0%; Score 114; DB 34; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.71e-41;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 atgatgatattataaaactttcagaaaagaattgcataaaataaacggtggtatgcca 60
QY 1 ATGATGATATTAAAAAATTTTCAGAAAAAGAAATTCAGAAAAATAAACGGTGTATGGCA 60
Db 61 ggaataagttcattttatcataagattaaacaaattttaccatcggttaa 114
QY 61 GGAATAAGTCTCTAATTTATTCATAGATTAAACAAATTTTACCCATCGTTAA 114
RESULT 2
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;
Query Match 26.3%; Score 30; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 8.34e-03;
Matches 0; Conservative 39; Mismatches 9; Indels 0; Gaps 0;
Db 12 svhsyvvvhhvshhshvhhvhhvsvvvvhhvhhvhhvhhvhhvhhv 59
QY 3 GATGATATTAAAAAATTTTCAGAAAAAGAAATTCAGAAAAATAAACGG 50
RESULT 3
ID V44650 standard; DNA; 91 BP.
AC V44650;
DT 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.
Query Match 26.3%; Score 30; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 8.34e-03;
Matches 0; Conservative 39; Mismatches 9; Indels 0; Gaps 0;
Db 12 svhsyvvvhhvshhshvhhvhhvsvvvvhhvhhvhhvhhvhhvhhv 59
QY 3 GATGATATTAAAAAATTTTCAGAAAAAGAAATTCAGAAAAATAAACGG 50

PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.
PI Cossens NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
DR WPI: 98-362770/31.
PT Human or mammalian origin of replication consensus sequences - for
PT inhibiting DNA replication, for controlling initiation of
PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1; Page 42; 54pp; English.
CC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorconsensus.
CC Administration of the consensus sequence is used to inhibit DNA replication
CC double stranded copy of the consensus) is used to insert into an
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned
CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular or
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC elements.
SQ Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;
Query Match 26.3%; Score 30; DB 46; Length 91;
Best Local Similarity 13.6%; Pred. No. 8.34e-03;
Matches 12; Conservative 47; Mismatches 29; Indels 0; Gaps 0;
Db 4 twaakrawwkkdvvvgakrwwkvvhrassacmdwkaaktwkggtwarrywkgkrm 63
Cp 103 TAAAAATTTGTTTAACTCTTATGAATAAAATAGAACTATTCTCTGCCATACCACCGTTTA 44
Db 64 wwtkawsdatakwkwwkdakwkmwrktt 91
Cp 43 TTTTGTGCAATCTTTTCTGAAAGTTT 16
RESULT 4
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;
Query Match 26.3%; Score 30; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 8.34e-03;

```

Matches      0; Conservative 40; Mismatches 10; Indels 0; Gaps 0;

Db 11 ssvhsyvvvshhshhvvhvshvsvvvhvhhvvhvhhvvhvhhvvsv 60
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 106 GGGTAAATTTGTTTAACTTATGAATAAATAGAACTATTTCTGCCC 57
    : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      5
ID N81164 standard; DNA; 204 BP.
AC N81164;
DE 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
    /tag= a
    /function=multiple cloning site
FT primer_bind 187..204
    /tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 25.4%; Score 29; DB 1; Length 204;
Best Local Similarity 8.0%; Pred. No. 2.10e-02;
Matches 8; Conservative 52; Mismatches 39; Indels 1; Gaps 1;

Db 91 thhyrrmbnvdyrdnrsdaawccyrsvkycnchddhvvbbvynvhn 150
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 8 TATTTAAATTTTATTCAGAAAAGAAATTCGAAAATAA-ACGGTGGTATGCGAGAAAT 66
    : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 151 cncccbnhchvbnhnnhrnwayrhdrrddvhecvc 190
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 67 AGTCTAAATTTTATTCATAAGATTAACAAATTTTATACC 106
    : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      6
ID N81164 standard; DNA; 204 BP.
AC N81164;
DE 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
    /tag= a
    /function=multiple cloning site
FT primer_bind 187..204
    /tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 25.4%; Score 29; DB 1; Length 204;
Best Local Similarity 8.0%; Pred. No. 2.10e-02;
Matches 8; Conservative 52; Mismatches 39; Indels 1; Gaps 1;

Db 91 thhyrrmbnvdyrdnrsdaawccyrsvkycnchddhvvbbvynvhn 150
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 8 TATTTAAATTTTATTCAGAAAAGAAATTCGAAAATAA-ACGGTGGTATGCGAGAAAT 66
    : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 151 cncccbnhchvbnhnnhrnwayrhdrrddvhecvc 190
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 67 AGTCTAAATTTTATTCATAAGATTAACAAATTTTATACC 106
    : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 25.4%; Score 29; DB 1; Length 204;
Best Local Similarity 10.1%; Pred. No. 2.10e-02;
Matches 10; Conservative 47; Mismatches 42; Indels 0; Gaps 0;

Db 87 mrttthyrmbnvdyrdnrsdaawccyrsvkycnchddhvvbbvynv 146
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 100 AAATTTGTTTATTCATGAATAAATAGAACTATTTCTGCGATACACGGTTATTT 41
    : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 147 nhnccbnhchvbnhnnhrnwayrhdrrddv 185
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 40 TTGCAATCTTTCTGAAAGTTTTTTAAATATCATCA 2
    : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      7
ID N50024 standard; DNA; 501 BP.
AC N50024;
DE 04-SEP-1991 (first entry)
DE Sequence encoding new modified human beta interferon polypeptides
DE IFN4 417.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501
    /tag= a
PN EP-163993-A.
PD 11-DEC-1985.
PF 17-MAY-1985; 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR ) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
DR WPI: 85-311944/50.
DR P-PSDB; P50023.
PT New modified human beta interferon polypeptide(s) - prepd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PT Claim 28; Chart 2b, page 33; 71pp; English.
CC Compared with interferon beta prepd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
CC See also P80575.
SQ Sequence 501 BP; 110 A; 32 C; 66 G; 81 T;

Query Match 24.6%; Score 28; DB 3; Length 501;
Best Local Similarity 44.6%; Pred. No. 5.26e-02;
Matches 25; Conservative 15; Mismatches 16; Indels 0; Gaps 0;

Db 366 baarmgdtaytggngmgdthybcaytaytbaatgcnaargattaywncatt 421

```

44 TAAACGGTGGTATGCCAGGAAATAGCTCTAATTTTATTCATAAGATTAAACAAATT 99

RESULT	8	
ID	V44650 standard; DNA; 91 BP.	
AC	V44650;	
DE	06-OCT-1998 (first entry)	
DE	Mammalian DNA replication origin consensus sequence, uniorsconsensus.	
DE	DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;	
KW	anti-gene; DNA replication inhibitor; shuttle vector construct creation;	
KW	gene therapy; ss.	
OS	Mammalia.	
OS	WO9827200-A2.	
PD	25-JUN-1998.	
PD	12-DEC-1997; CA0972.	
PR	21-MAY-1997; US-047322.	
PR	16-DEC-1996; US-033374.	
PA	(UYMC-) UNIV MCGILL.	
PA	COSSONS NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;	
WI	WPI; 98-362770/31.	
DR	Human or mammalian origin of replication consensus sequences - for	
PT	inhibiting DNA replication, for controlling initiation of	
PT	replication, maintaining circular plasmids and in assembly of human	
PT	artificial chromosomes	
PT	Claim 1; Page 42; 54pp; English.	
PS	This sequence represents a human or mammalian DNA replication origin	
CC	consensus sequences of the invention, designated uniorsconsensus.	
CC	Administration of the consensus sequence or an anti-gene (comprising a	
CC	double stranded copy of the consensus) is used to inhibit DNA replication	
CC	in vivo or in vitro. The consensus sequences can also be inserted into a	
CC	expression vector, used subsequently for in vitro transfection of	
CC	mammalian cells, to control initiation of DNA replication. They can also	
CC	be used to maintain circular plasmids that are capable of	
CC	semi-conservative replication in proliferating mammalian cells, or	
CC	inserted into mammalian or human artificial chromosome vectors for gene	
CC	therapy. Particularly, they are used to create shuttle vector constructs	
CC	for defining the essential mammalian elements required for maintenance of	
CC	chromosomal function. The consensus sequence can be combined with cloned	
CC	human telomeres and large centromeric blocks for assembly of human	
CC	artificial chromosomes and maintained as bacterial plasmids, circular or	
CC	linear, large or small yeast artificial chromosomes (YACs) or as episomal	
CC	elements.	
CC	Sequence. 91 BP: 15 A: 1 C: 4 G: 7 T;	

RESULT	9	
ID	N50023 standard; DNA; 501 BP.	
AC	N50023;	
DT	04-SEP-1991 (first entry)	
DE	Sequence encoding new modified human beta interferon polypeptides	
DE	IFN γ 416.	
DE	Antiviral; cell growth regulator; immune system regulator;	
KW	antiproliferative; ss.	
OS	Homo sapiens.	
FS	Key	Location/Qualifiers
FT	Cds	1..501
FT		/*tag= a
FT	EP-163993-A.	
PN	11-DEC-1985.	
PD	17-MAY-1985; 105750.	
PF		

PR 17-MAY-1984; GB-012564.
PA (SEAR) SEARLE G D & CO.
PI Bell LD, Boseley PG, Porter AG;
WPI; 85-311944/50.
DR P-PSDB; P50022.
PT New modified human beta interferon polypeptide(s) - prep'd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PS Claim 28; Chart 2a, page 32; 71pp; English.
CC Compared with interferon beta prep'd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. They are also easier to recover from incubation mixts.
SQ Sequence 501 BP; 107 A; 31 C; 69 G; 80 T;
Query Match 22.8%; Score 26; DB 3; Length 501;
Best Local Similarity 41.1%; Pred. No. 3,21e-01;
Waste-bac 23; Conservative 16; Mismatches 17; Indels 0; Gaps

RESULT	10	
ID	N50026 standard; DNA; 501 BP.	
AC	N50026;	
DE	4-SEP-1991 (first entry)	
DT	Sequence encoding new modified human beta interferon polypeptides	
DE	IFNX 430.	
DE	Antiviral; cell growth regulator; immune system regulator;	
KW	antiproliferative; ss.	
KW	Homo sapiens.	
OS		
Key	Location/Qualifiers	
FT	1..501	
ET	/*tag= a	
FP	P-163993-A.	
PD	11-DEC-1985.	
PF	17-MAY-1985; 105750.	
PR	17-MAY-1984; GB-012564.	
PA	(SEAR) SEARLE G D & CO.	
PI	Bell LD, Boseley PG, Porter AG;	
PI	WPI; 85-311944/50.	
DR	P-PSDB; P50025.	
DR	New modified human beta interferon polypeptide(s) - prepd. by	
PT	plasmid transformed bacteria, with improved antiviral,	
PT	anti-proliferative and immune regulating actions	
PT	Claim 28: Chart 2d, page 35; 71pp; English.	
CC	Compared with interferon beta prepd. by recombinant methods, the	
CC	INFS of the invention are more active and have different affinities	
CC	for cell surface receptors (allowing selective targeting); they	
CC	have higher therapeutic index; Improved stability against microbial	
CC	breakdown during synthesis; and better in vivo solubility and	
CC	stability. They are also easier to recover from incubation mixts.	
CC	501 Bp. 108 A. 70 G. 81 T.	

RESULT	11
ID	Q70467 standard; DNA; 114 BP.
AC	Q70467;
DT	05-APR-1995 (first entry)
DE	Generic DNA sequence to generate a random TSAR peptide library.
DE	TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW	TSAR; totally synthetic affinity reagent; synthetic; binding domain;

W P S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:38:38 1999; MasPar time 9.71 Seconds
Tabular output not generated. 831.838 Million cell updates/sec

Title: >US-09-068-507A-2
Description: (1-114) from US09068507A.seq
Perfect Score: 114
N.A. Sequence: 1 ATGATGATATTTAAAAACT.....AAATTTTACCCATCGTTAA 114
Comp: TACTACTATAATTTTGA.....TTTAAATGGTAGCAATT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 137068 seqs, 35432894 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 6.588; Variance 4.618; scale 1.427

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	27	23.7	7218	2	US-08-232-Sequence 14, Applicatio	1,41e-03
2	22	19.3	2183	1	US-07-993-Sequence 1, Applicatio	3,36e-01
3	22	19.3	3440	4	PCT-US91-0-Sequence 27, Applicati	3,36e-01
4	22	19.3	3440	2	US-08-471-Sequence 27, Applicati	3,36e-01
5	21	18.4	928	4	PCT-US96-0-Sequence 5, Applicatio	9,62e-01
6	21	18.4	1090	4	PCT-US96-0-Sequence 4, Applicatio	9,62e-01
7	21	18.4	1407	4	PCT-US96-0-Sequence 3, Applicatio	9,62e-01
8	21	18.4	1461	4	PCT-US96-0-Sequence 27, Applicati	9,62e-01
9	21	18.4	1791	4	PCT-US96-0-Sequence 10, Applicatio	9,62e-01
10	21	18.4	3291	4	PCT-US96-0-Sequence 1, Applicatio	9,62e-01
11	21	18.4	5595	1	US-07-635-Sequence 1, Applicatio	9,62e-01
12	21	18.4	5595	1	US-07-841-Sequence 1, Applicatio	9,62e-01
13	21	18.4	5648	4	PCT-US96-0-Sequence 1, Applicatio	9,62e-01
14	20	17.5	215	1	US-08-238-Sequence 5, Applicatio	2,71e+00
15	20	17.5	215	1	US-08-238-Sequence 5, Applicatio	2,71e+00
16	20	17.5	729	3	US-08-189-Sequence 9, Applicatio	2,71e+00
17	20	17.5	729	3	US-08-189-Sequence 26, Applicati	2,71e+00
18	20	17.5	965	3	US-08-388-Sequence 22, Applicati	2,71e+00
19	20	17.5	1670	3	US-08-188-Sequence 5, Applicatio	2,71e+00
20	20	17.5	1670	1	US-08-189-Sequence 10, Applicati	2,71e+00

21	20	17.5	1969	3	US-08-188-Sequence 3, Applicatio	2,71e+00
22	20	17.5	2687	1	US-08-149-Sequence 8, Applicatio	2,71e+00
23	20	17.5	2687	1	US-08-377-Sequence 8, Applicatio	2,71e+00
24	20	17.5	2873	1	US-08-149-Sequence 1, Applicatio	2,71e+00
25	20	17.5	2873	1	US-08-377-Sequence 1, Applicatio	2,71e+00
26	20	17.5	2913	3	US-08-188-Sequence 6, Applicatio	2,71e+00
27	20	17.5	3180	4	PCT-US96-0-Sequence 1, Applicatio	2,71e+00
28	20	17.5	3180	3	US-08-480-Sequence 1, Applicatio	2,71e+00
29	19	16.7	1145	5	5510472-1 Patent No. 5510472	7,47e+00
30	19	16.7	1288	3	US-08-440-Sequence 9, Applicatio	7,47e+00
31	19	16.7	2051	3	US-08-482-Sequence 2, Applicatio	7,47e+00
32	19	16.7	2975	2	US-08-368-Sequence 1, Applicatio	7,47e+00
33	19	16.7	3033	3	US-08-482-Sequence 1, Applicatio	7,47e+00
34	19	16.7	3240	2	US-08-368-Sequence 3, Applicatio	7,47e+00
35	19	16.7	3303	1	US-08-081-Sequence 3, Applicatio	7,47e+00
36	18	15.8	198	4	PCT-US92-1-Sequence 3, Applicatio	7,47e+00
37	18	15.8	660	1	US-07-991-Sequence 16, Applicati	2,02e+01
38	18	15.8	1511	1	US-07-991-Sequence 32, Applicati	2,02e+01
39	18	15.8	1511	2	US-08-107-Sequence 8, Applicatio	2,02e+01
40	18	15.8	2306	4	PCT-US96-0-Sequence 8, Applicatio	2,02e+01
41	18	15.8	2760	1	US-08-101-Sequence 3, Applicatio	2,02e+01
42	18	15.8	2760	1	US-08-101-Sequence 1, Applicatio	2,02e+01
43	18	15.8	4837	1	US-08-089-Sequence 3, Applicatio	2,02e+01
44	18	15.8	8457	1	US-07-991-Sequence 1, Applicatio	2,02e+01
45	18	15.8	9636	3	US-08-323-Sequence 1, Applicatio	2,02e+01

ALIGNMENTS

RESULT 1
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC xxxxxx
DT
DE Sequence 14, Application US/08232463
CC Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSES: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232.463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935.313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/POCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:

CC	TITLE OF INVENTION:	REST Protein and DNA
CC	NUMBER OF SEQUENCES:	29
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Dechert Price & Rhoads
CC	STREET:	P.O. Box 5218
CC	CITY:	Princeton
CC	STATE:	New Jersey
CC	COUNTRY:	USA
CC	ZIP:	08543-5218
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Diskette, 3.50 inch, 1.44 Mb storage
CC	COMPUTER:	IBM-compatible
CC	OPERATING SYSTEM:	DOS 5.0
CC	SOFTWARE:	WordPerfect
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	PCT/US96/03940
CC	FILING DATE:	March 23, 1995
CC	CLASSIFICATION:	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Allen Bloom
CC	REGISTRATION NUMBER:	29,135
CC	REFERENCE/DOCKET NUMBER:	31743-101
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(609) 520-3214
CC	TELEFAX:	(609) 520-3259
CC	INFORMATION FOR SEQ ID NO: 5:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	928 base pairs
CC	TYPE:	nucleic acid
CC	STRANDEDNESS:	double
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	cDNA to mRNA
CC	HYPOTHETICAL:	no
CC	ANTI-SENSE:	no
CC	ORIGINAL SOURCE:	
CC	ORGANISM:	Human
CC	CELL LINE:	HeLa
CC	IMMEDIATE SOURCE:	
CC	LIBRARY:	cDNA
CC	PUBLICATION INFORMATION:	
CC	AUTHORS:	Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC	AUTHORS:	Aral, Juan, Zheng, Yingcong, Boutros, Michael C
CC	AUTHORS:	Yelena M., Frohman, Michael A., Kraner, Susan D
CC	TITLE:	REST: A Mammalian Silencer Protein that Restricts
CC	TITLE:	Sodium Channel Gene Expression to Neurons
CC	JOURNAL:	Cell
CC	VOLUME:	80
CC	ISSUE:	
CC	PAGES:	
CC	DATE:	March 24, 1995
CC	RELEVANT RESIDUES IN SEQ ID NO:	5:FROM 1 TO 928
CC	SEQUENCE 928 BP;	363 A; 178 C; 182 G; 205 T; 0 OTHER.
	Query Match	18.4%; Score 21; DB 4; Length 928;
	Best Local Similarity	74.4%; Pred. No. 9,62e-01;
	Matches	32; Conservative 0; Mismatches 11; Indels 0;
Db	643	CAGAAATAGACAACAAAATAAAGGGGATGTGGCTGGAAA 685
QY	23	CAGAAAAAGAATTGCCAAAAATAAACGTCGTATGGCAGAAA 65
RESULT	6	
ID	PCT-US96--03940-4	STANDARD; DNA; UNC; 1090 BP.
ACC	xxxxxx	
DT		
DE	Sequence 4, Application PC/TUS9603940	
CC	Sequence 4, Application PC/TUS9603940	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Mandel, Gall, Chong, Jayhong A.
CC	TITLE OF INVENTION:	REST Protein and DNA
CC	NUMBER OF SEQUENCES:	29
CC	CORRESPONDENCE ADDRESS:	

CC	STATE: New Jersey
CC	COUNTRY: USA
CC	ZIP: 08543-5218
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC	COMPUTER: IBM-compatible
CC	OPERATING SYSTEM: DOS 5.0
CC	SOFTWARE: WordPerfect
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US96/03940
CC	FILING DATE: March 23, 1995
CC	CLASSIFICATION:
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Allen Bloom
CC	REGISTRATION NUMBER: 29,135
CC	REFERENCE/DOCKET NUMBER: 317743-101
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: (609) 520-3214
CC	TELEFAX: (609) 520-3259
CC	INFORMATION FOR SEQ ID NO: 3:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 1407 base pairs
CC	TYPE: nucleic acid
CC	STRANDEDNESS: double
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: cDNA to mRNA
CC	HYPOTHETICAL: no
CC	ANTI-SENSE: no
CC	ORIGINAL SOURCE:
CC	ORGANISM: Human
CC	CELL LINE: HeLa
CC	IMMEDIATE SOURCE:
CC	LIBRARY: cDNA
CC	PUBLICATION INFORMATION:
CC	AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC	AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC	AUTHORS: Yelena M., Frohman, Michael A., Krainer, Susan D., Mandel, Gail
CC	TITLE: REST: A Mammalian Silencer Protein that Restricts
CC	TITLE: Sodium Channel Gene Expression to Neurons
CC	JOURNAL: Cell
CC	VOLUME: 80
CC	ISSUE:
CC	PAGES:
CC	DATE: March 24, 1995
CC	RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 1407
CC	SEQUENCE 1407 BP; 515 A; 276 C; 306 G; 310 T; 0 OTHER.
CC	Query Match 18.4%; Score 21; DB 4; Length 1407;
CC	Best Local Similarity 74.4%; Pred. No. 9.62e-01;
CC	Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
DB	1122 CAGAAATAGACACACAAATAAATAAGGGGATGTGGCTGGAAA 1164
QY	23 CAGAAAAAGAATTGCAAAAAATAAACCGTGGTATGCGAGAAA 65
RESULT	8
ID	PCT-US96-03940-27 STANDARD; DNA; UNC; 1461 BP.
AC	xxxxxx
DT	
DE	Sequence 27, Application PC/TUS9603940
CC	Sequence 27, Application PC/TUS9603940
CC	GENERAL INFORMATION:
CC	APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC	TITLE OF INVENTION: REST Protein and DNA
CC	NUMBER OF SEQUENCES: 29
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Dechert Price & Rhoads
CC	STREET: P.O. Box 5218
CC	CITY: Princeton
CC	STATE: New Jersey
CC	COUNTRY: USA
CC	ZIP: 08543-5218

```

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 27:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1461 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna to mrna
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORIGINAL SOURCE:
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: CDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D., Mandel, Gail
CC TITLE: REST: A Mammalian Silencer Protein that Restricts
CC TITLE: Sodium Channel Gene Expression to Neurons
CC JOURNAL: Cell
CC VOLUME: 80
CC ISSUE:
CC DATE: March 24, 1995
CC PAGES:
CC RELEVANT RESIDUES IN SEQ ID NO: 26: FROM 1 TO 1461
CC SEQUENCE 1461 BP: 528 A; 288 C; 326 G; 319 T; 0 OTHER.

Query Match 18.4%; Score 21; DB 4; Length 1461;
Best Local Similarity 74.4%; Pred. No. 9.62e-01;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1217 CAGAAATAGACAAACAAAATAAAGGGGATGCGTGCGGAAA 1259
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 23 CAGAAAAGAAATTCGAAAAATAAACGCGTGTATGCGAGGAAA 65

RESULT 9
ID PCT-US96-03940-6 STANDARD; DNA; UNC; 1791 BP.
AC xxxxxx
DT
DE Sequence 6, Application PC/TUS9603940
CC Sequence 6, Application PC/TUS9603940
CC GENERAL INFORMATION:
CC APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC TITLE OF INVENTION: REST Protein and DNA
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: P.O. Box 5218
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible

```

```

CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1791 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna to mrna
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORIGINAL SOURCE:
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: CDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D., Mandel, Gail
CC TITLE: REST: A Mammalian Silencer Protein that Restricts
CC TITLE: Sodium Channel Gene Expression to Neurons
CC JOURNAL: Cell
CC VOLUME: 80
CC ISSUE:
CC DATE: March 24, 1995
CC PAGES:
CC RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 1791
CC SEQUENCE 1791 BP: 607 A; 395 C; 406 G; 383 T; 0 OTHER.

Query Match 18.4%; Score 21; DB 4; Length 1791;
Best Local Similarity 74.4%; Pred. No. 9.62e-01;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1512 CAGAAATAGACAAACAAAATAAAGGGGATGCGTGCGGAAA 1554
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 23 CAGAAAAGAAATTCGAAAAATAAACGCGTGTATGCGAGGAAA 65

RESULT 10
ID PCT-US96-03940-10 STANDARD; DNA; UNC; 3291 BP.
AC xxxxxx
DT
DE Sequence 10, Application PC/TUS9603940
CC Sequence 10, Application PC/TUS9603940
CC GENERAL INFORMATION:
CC APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC TITLE OF INVENTION: REST Protein and DNA
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: P.O. Box 5218
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:

```


CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: P.O. Box 5218
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC MEDIUM TYPE: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 5648 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA to mRNA
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORIGINAL SOURCE:
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: cDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos ,
CC AUTHORS: Toledo-Aral,
CC AUTHORS: Juan, Zheng, Yingcong, Boutros, Michael C
CC AUTHORS: M., Frohman, Michael A., Kraner, Susan D
CC TITLE: REST: A Mammalian Silencer Protein that Res
CC TITLE: Sodium Channel Gene Expression to Neurons
CC JOURNAL: Cell
CC VOLUME: 80
CC ISSUE:
CC PAGES:
CC DATE: March 24, 1995
CC RELEVANT RESIDUES IN SEQ ID NO: 1:FROM -1 TO 5648
CC SEQUENCE 5648 BP: 1747 A; 1098 C; 1294 G; 1507 T; 2 OTHER

	Query Match	18.4%	Score 21:	DB 4:	Length 5648:
	Best Local Similarity	74.4%	Pred. No. 9,62e-01:		
	Matches	32:	Conservative	0:	Mismatches 11: Indels
Db	1661	CAGAAATAGACAAACAAAAATAAAGGGGATGTGGCTGGAA	1703		
QY	23	CAGAAAAGGAATTGCAAAAAATAAACGGTGGTATGGCAGAA	65		

RESULT	14
ID	US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC	xxxxxx
DT	
DE	Sequence 5, Application US/08238163
CC	Sequence 5, Application US/08238163
CC	Patent No. 5369830
CC	GENERAL INFORMATION:
CC	APPLICANT: BENNETT, Alan
CC	APPLICANT: LABAVITCH, John M.
CC	APPLICANT: FOWELL, Ann
CC	APPLICANT: STORZ, Henrik
CC	TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL

M E S R A

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:47:40 1999; MasPar time 268.21 Seconds
Tabular output not generated. 995.949 Million cell updates/sec

Title: >US-09-068-507A-2
Description: (1-114) from US09068507A.seq
Perfect Score: 114
N.A. Sequence: 1 ATGATGATATTATAAAACTAAATTTTACCATCGTTAA 114
Comp: TACTACTATAAATTTTGA.....TTTAAATGGTAGCAATT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est11
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 9.510; Variance 3.329; scale 2.857

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	31	27.2	252	17	97SN1787 Rice Immature	3.63e-08
C 2	30	26.3	2275	20	AF034173 Human mRNA (T	2.71e-07
C 3	29	25.4	252	17	97SN1787 Rice Immature	1.97e-06
C 4	25	21.9	365	24	aa47e06.xl Stanley Fro	4.01e-03
C 5	25	21.9	392	24	aa47e06.xl Stanley Fro	4.01e-03
C 6	25	21.9	675	39	CPG0587A CpiOWAGDNL C	4.01e-03
C 7	24	21.1	356	40	AQ268382 RPC111-69M19.TJ R	2.45e-02
C 8	24	21.1	359	18	aa47e06.xl Stanley Fro	2.45e-02
C 9	24	21.1	423	25	AQ039222 Dictyostelium	2.45e-02
C 10	24	21.1	434	40	AQ268460 RPC111-74G21.TJ R	2.45e-02

11	24	21.1	435	24	AU033574	AU033574 Dictyostelium	2.45e-02
C 12	24	21.1	530	41	AQ344881	RPC111-114117.TJ R	2.45e-02
C 13	23	20.2	133	29	A5170972	to24603.xl NCI_CGAP_U	1.44e-01
C 14	23	20.2	252	21	AA998637	MCCE3650.MLF Schistos	1.44e-01
C 15	23	20.2	233	11	AA344668	EST38687 Embryo, 9 wee	1.44e-01
C 16	23	20.2	300	15	C59929	C59929 Yuji Kohara unp	1.44e-01
C 17	23	20.2	338	8	D32810	CELK021HYR Yuji Kohara	1.44e-01
C 18	23	20.2	349	27	A178233	tm50f12.xl NCI_CGAP_K	1.44e-01
C 19	23	20.2	352	22	A1039179	ox31e12.s1 Soares tota	1.44e-01
C 20	23	20.2	362	12	AA371866	EST83933 Parathyroid 9	1.44e-01
C 21	23	20.2	394	38	B55192	CIT-HSP-386F22.TR CIT-	1.44e-01
C 22	23	20.2	400	30	R50363	YJ59605.s1 Soares brea	1.44e-01
C 23	23	20.2	403	17	AA748479	ny01f04.s1 NCI_CGAP_G	1.44e-01
C 24	23	20.2	430	8	D20068	HUMG01040 Human promy	1.44e-01
C 25	23	20.2	441	8	D32326	CELK010G9R Yuji Kohara	1.44e-01
C 26	23	20.2	515	21	C92718	C92718 Dictyostelium d	1.44e-01
C 27	23	20.2	542	22	A1066149	TENU2540 T. cruzi eplm	1.44e-01
C 28	23	20.2	551	21	C92168	C92168 Dictyostelium d	1.44e-01
C 29	23	20.2	568	41	AQ372536	RPC111-113020.TJ R	1.44e-01
C 30	23	20.2	612	41	AQ375760	RPC111-150J22.TV R	1.44e-01
C 31	23	20.2	724	40	AQ269610	HS_2036.B1_G07_MR CIT	1.44e-01
C 32	23	20.2	759	40	AQ253262	HS_2046.B1_A08_MR CIT	1.44e-01
C 33	23	20.2	763	28	A1537158	PZ2_114_C01.r tumor2	1.44e-01
C 34	23	20.2	804	37	B12681	F27D1-Sp6.1 IGF Arabid	1.44e-01
C 35	23	20.2	943	37	B13335	F27D1-T7 IGF Arabidopsi	1.44e-01
C 36	22	19.3	362	31	R97401	YQ53B06.s1 Soares feta	8.04e-01
C 37	22	19.3	367	9	AA200306	mu58d10.r1 Soares feta	8.04e-01
C 38	22	19.3	396	38	AQ015142	CIT-HSP-231J34.TR CIT-	8.04e-01
C 39	22	19.3	398	31	H44951	y069d08.s1 Soares brea	8.04e-01
C 40	22	19.3	411	33	N68561	za15c07.s1 Soares feta	8.04e-01
C 41	22	19.3	468	33	N52522	yy52g10.s1 Soares feta	8.04e-01
C 42	22	19.3	468	27	A1447234	mg97h06.xl Soares mous	8.04e-01
C 43	22	19.3	528	42	AQ392828	CJTBI-E1-2544J24.TR CI	8.04e-01
C 44	22	19.3	632	24	AU030522	AU030522 Rice cDNA fro	8.04e-01
C 45	22	19.3	653	17	AA636365	GM07543.5prime GM Dros	8.04e-01

ALIGNMENTS

1
RESULT 1
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahn,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahn, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bnaam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .252
/organism="Oryza sativa"

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeunsun20@asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhna@mblserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.

```

FEATURES
source
location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pluscript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev stages="5 days after pollination"

```

```

BASE COUNT      5 a      21 c      12 g      35 t      179 others
ORIGIN

Query Match      25.4%; Score 29; DB 17; Length 252;
Best Local Similarity 12.7%; Pred. No. 1.97e-06;
Matches 14; Conservative 49; Mismatches 47; Indels 0; Gaps 0;

```

[illegible]

Db	76	DWHYWBVNTKVDGVNHTRCRWRBVTRMAHYHDYTNCBYNNNDYHMH	125
		::::: : : : : : : : : : : :	
Qv	63	AAATAGTCTAATTTTATTCAAGAGTAAACAAATTTTACCATCGTT	112

RESULT	4	AI267433	365 bp	mRNA	EST	17-NOV-1998
LOCUS		aq47e06.x1				
DEFINITION		Stanley Frontal NB pool 2 Homo sapiens cDNA clone IMAGE:2034082, mRNA sequence.				

DEFINITION	IMAGE: 2034082, mRNA sequence.
ACCESSION	AI267433
NID	q3886600
VERSION	AI267433.1
KEYWORDS	GT:3886600 EST.

VERSION	AI267433.1	GI:3886600
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

SOURCE	human.
ORGANISM	Homo sapiens

15-2-1

REFERENCE
AUTHORS
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE
JOURNAL
COMMENT
WashU-NCI human EST Project
Unpublished (1997)
On Oct 2, 1998 this sequence version replaced.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 334.
Location/Qualifiers
1. .365

FEATURES
source

/organism="Homo sapiens"
/note="Organ: brain; Vector: PCR2.1-TOPO (Invitrogen); Site: 1: ECKRI; Total RNA (purified with Trizol and DNaseI before use) was reverse transcribed using a modified oligo-dT primer containing RsaI and HindIII sites. Double-stranded cDNA was digested with RsaI, resulting in blunt ended cDNA of an average 0.1-2 kb in length. Digested cDNA was split into two sets, one used as is as the driver, the other set was split in half again and each half linked to a different adaptor (5'-TCGAGCGCGCCGCGCAGGT-3' or 5'-AGGCGTGTGGGAGGCGGT-3'), to be used as tester. Subtraction was performed using the Clontech PCR Select cDNA subtraction kit. Pool of two mentally normal males, ages 41 and 53 (S-124, S-141) subtracted by pool of two bipolar males, ages 45 and 50 (S-111, S-128). Tissues were obtained from the Stanley Neuropathology Consortium (www.stanleylab.org). Library constructed and subtracted by Dr. Nancy Johnston [(410) 614-3918, nij@welchlink.welch.jhu.edu]."
/db_xref="taxon:9606"
/map="21q"
/clone="IMAGE:2034082"
/clone_lib="Stanley Frontal NB pool 2"
/tissue_type="frontal lobe (see description)"
/lab_host="DH10B (phage-resistant)"
BASE COUNT 129 a 57 c 73 g 106 t
ORIGIN

Query Match 21.9%; Score 25; DB 24; Length 365;
Best Local Similarity 79.1%; Pred. No. 4.01e-03;
Matches 34; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 52 TTACATTTTGCAAATCTTTTTTTTAAATTTTAAATTTAT 94
||||| ||||||| ||||||| ||||||| ||||||| ||
Cp 46 TTATTTTTCGAATCTTTTCTGAAAGTTTTTAAATATCAT 4

RESULT 5
LOCUS
DEFINITION
a46e06.x1 Stanley Frontal NB pool 2 Homo sapiens cDNA clone
IMAGE:2033986, mRNA sequence.
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE
JOURNAL
COMMENT
WashU-NCI human EST Project
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2150725.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 378.
Location/Qualifiers
1. .392

FEATURES
source

/organism="Homo sapiens"
/note="Organ: brain; Vector: PCR2.1-TOPO (Invitrogen); Site: 1: ECKRI; Total RNA (purified with Trizol and DNaseI before use) was reverse transcribed using a modified oligo-dT primer containing RsaI and HindIII sites. Double-stranded cDNA was digested with RsaI, resulting in blunt ended cDNA of an average 0.1-2 kb in length. Digested cDNA was split into two sets, one used as is as the driver, the other set was split in half again and each half linked to a different adaptor (5'-TCGAGCGCGCCGCGCAGGT-3' or 5'-AGGCGTGTGGGAGGCGGT-3'), to be used as tester. Subtraction was performed using the Clontech PCR Select cDNA subtraction kit. Pool of two mentally normal males, ages 41 and 53 (S-124, S-141) subtracted by pool of two bipolar males, ages 45 and 50 (S-111, S-128). Tissues were obtained from the Stanley Neuropathology Consortium (www.stanleylab.org). Library constructed and subtracted by Dr. Nancy Johnston [(410) 614-3918, nij@welchlink.welch.jhu.edu]."
/db_xref="taxon:9606"
/map="6; 21q"
/clone="IMAGE:2033986"
/clone_lib="Stanley Frontal NB pool 2"
/tissue_type="frontal lobe (see description)"
/lab_host="DH10B (phage-resistant)"
BASE COUNT 138 a 58 c 77 g 119 t
ORIGIN

Query Match 21.9%; Score 25; DB 24; Length 392;
Best Local Similarity 79.1%; Pred. No. 4.01e-03;
Matches 34; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 52 TTACATTTTGCAAATCTTTTTTTTAAATTTTAAATTTAT 94
||||| ||||||| ||||||| ||||||| ||||||| ||
Cp 46 TTATTTTTCGAATCTTTTCTGAAAGTTTTTAAATATCAT 4

RESULT 6
LOCUS
DEFINITION
CpG0587A CpIOWAGDNA1 Cryptosporidium parvum genomic, genomic survey sequence.
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
Cryptosporidium parvum.
Cryptosporidium parvum
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
Cryptosporidiidae; Cryptosporidium.
REFERENCE
AUTHORS
Strong, W.B. and Nelson, R.G.

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:10090"
/clone="IMAGE:1225355"
/clone_lib="Soares 2NBMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"

BASE COUNT 107 a 56 c 65 g 131 t

Query Match 21.1%; Score 24; DB 18; Length 359;

Best Local Similarity 76.1%; Pred. No. 2.45e-02; Mismatches 0; Gaps 0;

Db 7 GATGATAATAAATAATTCACAAAGAGCAATGAAAAAATAAAC 52

Oy 3 GATGATATTTAAAAAACTTTCAGAAAAAGAAATTCAAAAAATAAAC 48

LOCUS AU039222 423 bp mRNA EST 29-MAR-1999
DEFINITION AU039222 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
ACCESSION AU039222
NID 94008203
VERSION AU039222.1 GI:4008203

KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 423)
Eukaryota: Dictyosteliida; Dictyostelium.

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takenoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227

COMMENT On Apr 14, 1993 this sequence version replaced gi:693192.
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES Location/Qualifiers
source 1..423
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLH180"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 173 a 57 c 44 g 149 t
ORIGIN
Query Match 21.1%; Score 24; DB 25; Length 423;
Best Local Similarity 71.4%; Pred. No. 2.45e-02;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 359 ATATAATAATAAATACTTCAACATAATTAGTATAATAATAAAGTAGTAT 414

Oy 1 ATGATGATATTTAAAAAACTTTCAGAAAAAGAAATTCAGAAAAAATAAACGGTGGTAT 56

RESULT 10

LOCUS

DEFINITION AU039222 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

ACCESSION AU039222

NID 94008203

VERSION AU039222.1 GI:4008203

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 423)

Eukaryota: Dictyosteliida; Dictyostelium.

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takenoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

TITLE The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development

JOURNAL DNA Res. 5 (6), 335-340 (1998)

COMMENT On Apr 14, 1993 this sequence version replaced gi:693192.

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.

FEATURES Location/Qualifiers

source 1..423

/organism="Dictyostelium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SLH180"

/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

/dev_stage="slug"

BASE COUNT 173 a 57 c 44 g 149 t

ORIGIN

Query Match 21.1%; Score 24; DB 25; Length 423;

Best Local Similarity 71.4%; Pred. No. 2.45e-02;

Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 359 ATATAATAATAAATACTTCAACATAATTAGTATAATAATAAAGTAGTAT 414

Oy 1 ATGATGATATTTAAAAAACTTTCAGAAAAAGAAATTCAGAAAAAATAAACGGTGGTAT 56

RESULT 10

LOCUS AU033574 435 bp mRNA EST 28-OCT-1998

DEFINITION AU033574 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

ACCESSION AU033574

NID 93798998

VERSION AU033574.1 GI:3798998

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 435)

Eukaryota: Dictyosteliida; Dictyostelium.

AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.

TITLE Developmental cDNA in Dictyostelium discoideum

JOURNAL Unpublished (1997)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1798828.

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.

FEATURES Location/Qualifiers

source 1..434

/organism="Homo sapiens"

/note="Vector: pACE3.6; Site_1: EcoRI; Site_2: EcoRI;"

/db_xref="GDB:7528196"

/db_xref="taxon:9606"

/clone="R-74G21"

/clone_lib="RPC11"

/sex="Male"

/cell_type="Lymphocytes"

BASE COUNT 147 a 74 c 89 g 124 t

ORIGIN

Query Match 21.1%; Score 24; DB 40; Length 434;

Best Local Similarity 76.1%; Pred. No. 2.45e-02;

Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 388 TGGATAAAGATTAGATTAGATGATGAATTAATAATTAAGAAATATCTC 433

Cp 107 TGGGTAAAAATTTGTTTAATCTTATGATAATAATAATTAAGACTATTC 62

RESULT 11

LOCUS AU033574 435 bp mRNA EST 28-OCT-1998

DEFINITION AU033574 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

ACCESSION AU033574

NID 93798998

VERSION AU033574.1 GI:3798998

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 435)

Eukaryota: Dictyosteliida; Dictyostelium.

AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.

TITLE Developmental cDNA in Dictyostelium discoideum

JOURNAL Unpublished (1997)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1798828.

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan

Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.

FEATURES Location/Qualifiers

source 1..434

/organism="Homo sapiens"

/note="Vector: pACE3.6; Site_1: EcoRI; Site_2: EcoRI;"

/db_xref="GDB:7528196"

/db_xref="taxon:9606"

/clone="R-74G21"

/clone_lib="RPC11"

/sex="Male"

/cell_type="Lymphocytes"

BASE COUNT 147 a 74 c 89 g 124 t

ORIGIN

Query Match 21.1%; Score 24; DB 40; Length 434;

Best Local Similarity 76.1%; Pred. No. 2.45e-02;

Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 388 TGGATAAAGATTAGATTAGATGATGAATTAATAATTAAGAAATATCTC 433

Cp 107 TGGGTAAAAATTTGTTTAATCTTATGATAATAATAATTAAGACTATTC 62

RESULT 11

ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 252)
AUTHORS Santos T.M., Azevedo, V., Marotta, G.B., Santos, R.L., Fonseca, S.J.,
Ortega, J.M., Rabelo, E.M.L., Saber, M., Abdel-Hamid, H., Ridgers, I.L.,
Johnston, D.A., Fernandez, M., Rollinson, D., Franco, G.R. and
Pena, S.D.J.
TITLE Analysis of the Gene expression profile in Schistosoma mansoni
cercariae using EST approach
JOURNAL Unpublished (1998)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043380.
Contact: Santos, T.M. and Pena, S.D.J.
Laboratorio de Genetica-Bioquimica, Departamento de Bioquimica e
Imunologia
Instituto de Ciencias Biologicas, Universidade Federal de Minas
Gerais
Avenida Antonio Carlos 6627, Belo Horizonte, MG, Brazil, 31270-010
Tel: (5531)4415611
Fax: (5531)4415409
Email: santostm@monoc.icb.ufmg.br
Seq primer: M13 Forward
FEATURES
source
1..252
Location/Qualifiers
/organism="Schistosoma mansoni"
/strain="NMRI"
/note="Vector: pBluescript SK; Site_1: XhoI; Site_2:
ECORI; mRNA was extracted from cercariae and the library
was constructed and excised according to the
manufacturer's instructions (Uni-zap XR vector,
Stratagene)"
/db_xref="taxon:6183"
/clone="SM3650C"
/clone_lib="Schistosoma mansoni, cercariae"
/dev_stage="Larvae"
/lab_host="DH5alpha"
BASE COUNT 97 a 60 c 15 g 80 t
ORIGIN
Query Match 20.2%; Score 23; DB 21; Length 252;
Best Local Similarity 71.7%; Pred. No. 1.44e-01;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Db 161 CCACCAGGAATTTTTCATTAATTTTCTATATTCATCAT 213
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 53 CCACCCTTTATTTTGCATCTTTTCTGAAGCTTTTAAATATCATCAT 1
RESULT 15
LOCUS AA334468 293 bp mRNA EST 21-APR-1997
DEFINITION EST38687 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA334468
NID 91986712
VERSION AA334468.1 GI:1986712
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 293)
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,

He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H.,
Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
Fraser, C.M. and Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
On Apr 14, 1993 this sequence version replaced gi:716849.
Other_ESTs: THC178342
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018659056
Fax: 3018659423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tldb/hgi/hgi.html>)
Seq primer: M13 Reverse
FEATURES
Location/Qualifiers
1..293
/organism="Homo sapiens"
/note="Vector: pBluescript SK; Site_1: EcorI; Site_2:
XhoI"
/db_xref="ATCC (inhost):136145"
/db_xref="taxon:9606"
/map="10"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
BASE COUNT 79 a 51 c 46 g 116 t 1 others
ORIGIN
Query Match 20.2%; Score 23; DB 11; Length 293;
Best Local Similarity 76.7%; Pred. No. 1.44e-01;
Matches 33; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Db 234 ATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 276
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4 ATGATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 46
Search completed: Wed Sep 15 10:52:51 1999
Job time : 311 secs.

